Wiley Handbooks in Financial Engineering and Econometrics

ADVANCES IN HEAVY TAILED RISK MODELING

A HANDBOOK OF OPERATIONAL RISK

Gareth W. Peters
Pavel V. Shevchenko

WILEY

Advances in Heavy Tailed Risk Modeling

A Handbook of Operational Risk

Advances in Heavy Tailed Risk Modeling

A Handbook of Operational Risk

Gareth W. Peters

Pavel V. Shevchenko

WILEY

Copyright © 2015 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Peters, Gareth W., 1978-
Advances in heavy tailed risk modeling : a handbook of operational risk / Gareth W. Peters, Department of Statistical Science, University College of London, London, United Kingdom, Pavel V. Shevchenko., Division of Computational Informatics, The Commonwealth Scientific and Industrial Research Organization, Sydney, Australia.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-90953-9 (hardback)
1. Risk management. 2. Operational risk. I. Shevchenko, Pavel V. II. Title.
HD61.P477 2014
658.15′5–dc23

2014015418

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Gareth W. Peters

This is dedicated to three very inspirational women in my life: Chen Mei–Peters, my mother Laraine Peters and Youxiang Wu; your support, encouragement and patience has made this possible. Mum, you instilled in me the qualities of scientific inquiry, the importance of questioning ideas and scientific rigour. This is especially for my dear Chen who bore witness to all the weekends in the library, the late nights reading papers and the ups and downs of toiling with mathematical proofs across many continents over the past few years.

Pavel V. Shevchenko

To my dear wife Elena

Embarking upon writing this book has proven to be an adventure through the landscape of ideas. Bringing forth feelings of adventure analogous to those that must have stimulated explorers such as Columbus to voyage to new lands.

In the depth of winter, I finally learned that within me there lay an invincible summer.
Albert Camus.

Contents in Brief

1 Motivation for Heavy-Tailed Models 1

2 Fundamentals of Extreme Value Theory for OpRisk 17

3 Heavy-Tailed Model Class Characterizations for LDA 105

4 Flexible Heavy-Tailed Severity Models: α-Stable Family 139

5 Flexible Heavy-Tailed Severity Models: Tempered Stable and Quantile Transforms 227

6 Families of Closed-Form Single Risk LDA Models 279

7 Single Risk Closed-Form Approximations of Asymptotic Tail Behaviour 353

8 Single Loss Closed-Form Approximations of Risk Measures 433

9 Recursions for Distributions of LDA Models 517

A Miscellaneous Definitions and List of Distributions 587

Contents

Preface xix

Acronyms xxi

Symbols xxiii

List of Distributions xxv

1 Motivation for Heavy-Tailed Models 1

1.1 Structure of the Book 1

1.2 Dominance of the Heaviest Tail Risks 3

1.3 Empirical Analysis Justifying Heavy-Tailed Loss Models in OpRisk 6

1.4 Motivating Parametric, Spliced and Non-Parametric Severity Models 9

1.5 Creating Flexible Heavy-Tailed Models via Splicing 11

2 Fundamentals of Extreme Value Theory for OpRisk 17

2.1 Introduction 17

2.2 Historical Perspective on EVT and Risk 18

2.3 Theoretical Properties of Univariate EVT–Block Maxima and the GEV Family 20

2.4 Generalized Extreme Value Loss Distributional Approach (GEV-LDA) 40

2.4.1 Statistical Considerations for Applicability of the GEV Model 40

2.4.2 Various Statistical Estimation Procedures for the GEV Model Parameters in OpRisk Settings 42

2.4.3 GEV Sub-Family Approaches in OpRisk LDA Modeling 54

2.4.4 Properties of the Frechet–Pareto Family of Severity Models 54

2.4.5 Single Risk LDA Poisson-Generalized Pareto Family 55
2.4.6 Single Risk LDA Poisson-Burr Family 60
2.4.7 Properties of the Gumbel family of Severity Models 65
2.4.8 Single Risk LDA Poisson-LogNormal Family 65
2.4.9 Single Risk LDA Poisson-Benktander II Models 68
2.5 Theoretical Properties of Univariate EVT–Threshold Exceedances 72
2.5.1 Understanding the Distribution of Threshold Exceedances 74
2.6 Estimation Under the Peaks Over Threshold Approach via the Generalized Pareto Distribution 85
2.6.1 Maximum-Likelihood Estimation Under the GPD Model 87
2.6.2 Comments on Probability-Weighted Method of Moments Estimation Under the GPD Model 93
2.6.3 Robust Estimators of the GPD Model Parameters 95
2.6.4 EVT—Random Number of Losses 101

3 Heavy-Tailed Model Class Characterizations for LDA 105
3.1 Landau Notations for OpRisk Asymptotics: Big and Little 'Oh' 106
3.2 Introduction to the Sub-Exponential Family of Heavy-Tailed Models 113
3.3 Introduction to the Regular and Slow Variation Families of Heavy-Tailed Models 121
3.4 Alternative Classifications of Heavy-Tailed Models and Tail Variation 129
3.5 Extended Regular Variation and Matuszewska Indices for Heavy-Tailed Models 135

4 Flexible Heavy-Tailed Severity Models: α-Stable Family 139
4.1 Infinitely Divisible and Self-Decomposable Loss Random Variables 140
4.1.1 Basic Properties of Characteristic Functions 140
4.1.2 Divisibility and Self-Decomposability of Loss Random Variables 143
4.2 Characterizing Heavy-Tailed α-Stable Severity Models 148
4.2.1 Characterisations of α-Stable Severity Models via the Domain of Attraction 152
4.3 Deriving the Properties and Characterizations of the α-Stable Severity Models 156
4.3.1 Unimodality of α-Stable Severity Models 158
4.3.2 Relationship between $\mathbb{L}$ Class and α-Stable Distributions 160
4.3.3 Fundamentals of Obtaining the α-Stable Characteristic Function 163
4.3.4 From Lévy–Khinchin's Canonical Representation to the α-Stable Characteristic Function Parameterizations 167

4.4 Popular Parameterizations of the α-Stable Severity Model Characteristic Functions 171
4.4.1 Univariate α-Stable Parameterizations of Zolotarev A, M, B, W, C and E Types 172
4.4.2 Univariate α-Stable Parameterizations of Nolan S0 and S1 178
4.5 Density Representations of α-Stable Severity Models 181
4.5.1 Basics of Moving from a Characteristic Function to a Distribution or Density 182
4.5.2 Density Approximation Approach 1: Quadrature Integration via Transformation and Clenshaw–Curtis Discrete Cosine Transform Quadrature 187
4.5.3 Density Approximation Approach 2: Adaptive Quadrature Integration via Fast Fourier Transform (Midpoint Rule) and Bergstrom Series Tail Expansion 189
4.5.4 Density Approximation Approach 3: Truncated Polynomial Series Expansions 196
4.5.5 Density Approximation Approach 4: Reparameterization 198
4.5.6 Density Approximation Approach 5: Infinite Series Expansion Density and Distribution Representations 200
4.6 Distribution Representations of α-Stable Severity Models 207
4.6.1 Quadrature Approximations for Distribution Representations of α-Stable Severity Models 208
4.6.2 Convergent Series Representations of the Distribution for α-Stable Severity Models 209
4.7 Quantile Function Representations and Loss Simulation for α-Stable Severity Models 210
4.7.1 Approximating the Quantile Function of Stable Loss Random Variables 210
4.7.2 Sampling Realizations of Stable Loss Random Variables 214
4.8 Parameter Estimation in an α-Stable Severity Model 215
4.8.1 McCulloch's Quantile-Based α-Stable Severity Model Estimators 216
4.8.2 Zolotarev's Transformation to W-Class-Based α-stable Severity Model Estimators 217
4.8.3 Press's Method-of-Moments-Based α-stable Severity Model Estimators 218
4.9 Location of the Most Probable Loss Amount for Stable Severity Models 219
4.10 Asymptotic Tail Properties of α-Stable Severity Models and Rates of Convergence to Paretian Laws 220

5 Flexible Heavy-Tailed Severity Models: Tempered Stable and Quantile Transforms 227
5.1 Tempered and Generalized Tempered Stable Severity Models 227

5.1.1 Understanding the Concept of Tempering Stable Severity Models 228
5.1.2 Families and Representations of Tempering in Stable Severity Models 231
5.1.3 Density of the Tempered Stable Severity Model 241
5.1.4 Properties of Tempered Stable Severity Models 243
5.1.5 Parameter Estimation of Loss Random Variables from a Tempered Stable Severity Model 246
5.1.6 Simulation of Loss Random Variables from a Tempered Stable Severity Model 248
5.1.7 Tail Behaviour of the Tempered Stable Severity Model 252
5.2 Quantile Function Heavy-Tailed Severity Models 253
5.2.1 g-and-h Severity Model Family in OpRisk 257
5.2.2 Tail Properties of the g-and-h, g, h and h-h Severity in OpRisk 268
5.2.3 Parameter Estimation for the g-and-h Severity in OpRisk 270
5.2.4 Bayesian Models for the g-and-h Severity in OpRisk 273

6 Families of Closed-Form Single Risk LDA Models 279
6.1 Motivating the Consideration of Closed-Form Models in LDA Frameworks 279
6.2 Formal Characterization of Closed-Form LDA Models: Convolutional Semi-Groups and Doubly Infinitely Divisible Processes 281
6.2.1 Basic Properties of Convolution Operators and Semi-Groups for Distribution and Density Functions 282
6.2.2 Domain of Attraction of Lévy Processes: Stable and Tweedie Convergence 303
6.3 Practical Closed-Form Characterization of Families of LDA Models for Light-Tailed Severities 309
6.3.1 General Properties of Exponential Dispersion and Poisson-Tweedie Models for LDA Structures 309
6.4 Sub-Exponential Families of LDA Models 321
6.4.1 Properties of Discrete Exponential Dispersion Models 322
6.4.2 Closed-Form LDA Models for Large Loss Number Processes 326
6.4.3 Closed-Form LDA Models for the α-Stable Severity Family 333
6.4.4 Closed-Form LDA Models for the Tempered α-Stable Severity Family 349

7 Single Risk Closed-Form Approximations of Asymptotic Tail Behaviour 353

7.1 Tail Asymptotics for Partial Sums and Heavy-Tailed Severity Models 356
7.1.1 Partial Sum Tail Asymptotics with Heavy-Tailed Severity Models: Finite Number of Annual Losses $N = n$ 357
7.1.2 Partial Sum Tail Asymptotics with Heavy-Tailed Severity Models: Large Numbers of Loss Events 362
7.2 Asymptotics for LDA Models: Compound Processes 367
7.2.1 Asymptotics for LDA Models Light Frequency and Light Severity Tails: SaddlePoint Tail Approximations 368
7.3 Asymptotics for LDA Models Dominated by Frequency Distribution Tails 372
7.3.1 Heavy-Tailed Frequency Distribution and LDA Tail Asymptotics (Frechet Domain of Attraction) 374
7.3.2 Heavy-Tailed Frequency Distribution and LDA Tail Asymptotics (Gumbel Domain of Attraction) 375
7.4 First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses 376
7.4.1 First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: General Sub-exponential Severity Model Results 377
7.4.2 First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Regular and O-Regularly Varying Severity Model Results 380
7.4.3 Remainder Analysis: First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models 385
7.4.4 Summary: First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models 388
7.5 Refinements and Second-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses 389
7.6 Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Dependent Losses 393
7.6.1 Severity Dependence Structures that Do Not Affect LDA Model Tail Asymptotics: Stochastic Bounds 402
7.6.2 Severity Dependence Structures that Do Not Affect LDA Model Tail Asymptotics: Sub-exponential, Partial Sums and Compound Processes 405

7.6.3 Severity Dependence Structures that Do Not Affect LDA Model Tail Asymptotics: Consistent Variation 410
7.6.4 Dependent Severity Models: Partial Sums and Compound Process Second-Order Tail Asymptotics 412
7.7 Third-order and Higher Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses 414
7.7.1 Background Understanding on Higher Order Tail Decomposition Approaches 414
7.7.2 Decomposition Approach 1: Higher Order Tail Approximation Variants 415
7.7.3 Decomposition Approach 2: Higher Order Tail Approximations 426
7.7.4 Explicit Expressions for Higher Order Recursive Tail Decompositions Under Different Assumptions on Severity Distribution Behaviour 430

8 Single Loss Closed-Form Approximations of Risk Measures 433

8.1 Summary of Chapter Key Results on Single-Loss Risk Measure Approximation (SLA) 433
8.2 Development of Capital Accords and the Motivation for SLAs 436
8.3 Examples of Closed-Form Quantile and Conditional Tail Expectation Functions for OpRisk Severity Models 440
8.3.1 Exponential Dispersion Family Loss Models 441
8.3.2 g-and-h Distribution Family Loss Models 445
8.3.3 Extended GPD: the Asymmetric Power Family Loss Models 446
8.4 Non-Parametric Estimators for Quantile and Conditional Tail Expectation Functions 448
8.5 First- and Second-Order SLA of the VaR for OpRisk LDA Models 451
8.5.1 Second-Order Refinements of the SLA VaR for Heavy-Tailed LDA Models 457
8.6 EVT-Based Penultimate SLA 468
8.7 Motivation for Expected Shortfall and Spectral Risk Measures 475
8.8 First- and Second-Order Approximation of Expected Shortfall and Spectral Risk Measure 478
8.8.1 Understanding the First-Order SLA for ES for Regularly Varying Loss Models 481
8.8.2 Second-Order SLA for Expected Shortfall for Regularly Varying Loss Models 485

8.8.3 Empirical Process and EVT Approximations of Expected Shortfall 488
8.8.4 SLA for Spectral Risk Measures 492
8.9 Assessing the Accuracy and Sensitivity of the Univariate SLA 496
8.9.1 Understanding the Impact of Parameter Estimation Error on a SLA 498
8.9.2 Understanding the SLA Error 502
8.10 Infinite Mean-Tempered Tail Conditional Expectation Risk Measure Approximations 503

9 Recursions for Distributions of LDA Models 517
9.1 Introduction 517
9.2 Discretization Methods for Severity Distribution 519
9.2.1 Discretization Method 1: Rounding 520
9.2.2 Discretization Method 2: Localized Moment Matching 522
9.2.3 Discretization Method 3: Lloyd's Algorithm 524
9.2.4 Discretization Method 4: Minimizing Kolmogorov Statistic 524
9.3 Classes of Discrete Distributions: Discrete Infinite Divisibility and Discrete Heavy Tails 525
9.4 Discretization Errors and Extrapolation Methods 533
9.5 Recursions for Convolutions (Partial Sums) with Discretized Severity Distributions (Fixed n) 535
9.5.1 De Pril Transforms for n-Fold Convolutions (Partial Sums) with Discretized Severity Distributions 537
9.5.2 De Pril's First Method 538
9.5.3 De Pril's Second Method 539
9.5.4 De Pril Transforms and Convolutions of Infinitely Divisible Distributions 540
9.5.5 Recursions for n-Fold Convolutions (Partial Sum) Distribution Tails with Discretized Severity 542
9.6 Estimating Higher Order Tail Approximations for Convolutions with Continuous Severity Distributions (Fixed n) 544
9.6.1 Approximation Stages to be Studied 547
9.7 Sequential Monte Carlo Sampler Methodology and Components 550
9.7.1 Choice of Mutation Kernel and Backward Kernel 553
9.7.2 Incorporating Partial Rejection Control into SMC Samplers 556

9.8 Multi-Level Sequential Monte Carlo Samplers for Higher Order Tail Expansions and Continuous Severity Distributions (Fixed n) 560
9.8.1 Key Components of Multi-Level SMC Samplers 562
9.9 Recursions for Compound Process Distributions and Tails with Discretized Severity Distribution (Random N) 565
9.9.1 Panjer Recursions for Compound Distributions with Discretized Severity Distributions 566
9.9.2 Alternatives to Panjer Recursions: Recursions for Compound Distributions with Discretized Severity Distributions 573
9.9.3 Higher Order Recursions for Discretized Severity Distributions in Compound LDA Models 575
9.9.4 Recursions for Discretized Severity Distributions in Compound Mixed Poisson LDA Models 577
9.10 Continuous Versions of the Panjer Recursion 581
9.10.1 The Panjer Recursion via Volterra Integral Equations of the Second Kind 581
9.10.2 Importance Sampling Solutions to the Continuous Panjer Recursion 583

A Miscellaneous Definitions and List of Distributions 587

A.1 Indicator Function, 587
A.2 Gamma Function, 587
A.3 Discrete Distributions, 587
A.3.1 Poisson Distribution, 587
A.3.2 Binomial Distribution, 588
A.3.3 Negative Binomial Distribution, 588
A.3.4 Doubly Stochastic Poisson Process (Cox Process), 589
A.4 Continuous Distributions, 589
A.4.1 Uniform Distribution, 589
A.4.2 Normal (Gaussian) Distribution, 590
A.4.3 Inverse Gaussian Distribution, 590
A.4.4 LogNormal Distribution, 591
A.4.5 Student's t-Distribution, 591
A.4.6 Gamma Distribution, 591
A.4.7 Weibull Distribution, 592
A.4.8 Inverse Chi-Squared Distribution, 592
A.4.9 Pareto Distribution (One Parameter), 592
A.4.10 Pareto Distribution (Two Parameter), 593
A.4.11 Generalized Pareto Distribution, 593

A.4.12 Beta Distribution, 594
A.4.13 Generalized Inverse Gaussian Distribution, 594
A.4.14 *d*-Variate Normal Distribution, 595
A.4.15 *d*-Variate *t*-Distribution, 595

REFERENCES 597

INDEX 623

Preface

This book covers key mathematical and statistical aspects of the quantitative modeling of heavy tailed loss processes in operational risk (OpRisk) and insurance settings. OpRisk has been through significant changes in the past few years with increased regulatory pressure for more comprehensive frameworks. Nowadays, every mid-sized and larger financial institution across the planet would have an OpRisk department. Despite the growing awareness and understanding of the importance of OpRisk modeling throughout the banking and insurance industry there is yet to be a convergence to a standardization of the modeling frameworks for this new area of risk management. In fact to date the majority of general texts on this topic of OpRisk have tended to cover basic topics of modeling that are typically standard in the majority of risk management disciplines. We believe that this is where the combination of the two books *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk* (Cruz, Peters and Shevchenko, 2015) and the companion book *Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk* will play an important role in better understanding specific details of risk modeling directly aimed to specifically capture fundamental and core features specific to OpRisk loss processes.

These two texts form a sequence of books which provide a detailed and comprehensive guide to the state of the art OpRisk modeling approaches. In particular, this second book on heavy tailed modeling provides one of the few detailed texts which is aimed to be accessible to both practitioners and graduate students with quantitative background to understand the significance of heavy tailed modeling in risk and insurance, particularly in the setting of OpRisk. It covers a range of modeling frameworks from general concepts of heavy tailed loss processes, to extreme value theory, how dependence plays a role in joint heavy tailed models, risk measures and capital estimation behaviors in the presence of heavy tailed loss processes and finishes with simulation and estimation methods that can be implemented in practice. This second book on heavy tailed modeling is targetted at a PhD or advanced graduate level quantitative course in OpRisk and insurance and is suitable for quantitative analysts working in OpRisk and insurance wishing to understand more fundamental properties of heavy tailed modeling that is directly relevant to practice. This is where the *Advances in Heavy-Tailed Risk Modeling: A Handbook of Operational Risk* can add value to the industry. In particular, by providing a clear and detailed coverage of modeling for heavy tailed OpRisk losses from both a rigorous mathematical as well as a statistical perspective.

More specifically, this book covers advanced topics on risk modeling in high consequence low frequency loss processes. This includes splice loss models and motivation for heavy tailed risk models. The key aspects of extreme value theory and their development in loss distributional approach modeling are considered. Classification and understanding of different classes of heavy tailed risk process models is discussed; this leads to topics on heavy tailed closed-form

loss distribution approach models and flexible heavy tailed risk models such as α-stable, tempered stable, g-and-h, GB2 and Tukey quantile transform based models. The remainder of the chapters covers advanced topics on risk measures and asymptotics for heavy tailed compound process models. Then the final chapter covers advanced topics including forming links between actuarial compound process recursions and Monte Carlo numerical solutions for capital risk measure estimations.

The book is primarily developed for advanced risk management practitioners and quantitative analysts. In addition, it is suitable as a core reference for an advanced mathematical or statistical risk management masters course or a PhD research course on risk management and asymptotics.

As mentioned, this book is a companion book of *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk* (Cruz, Peters and Shevchenko, 2015). The latter covers fundamentals of the building blocks of OpRisk management and measurement related to Basel II/III regulation, modeling dependence, estimation of risk models and the four-data elements (internal data, external data, scenario analysis and business environment and internal control factors) that need to be used in the OpRisk framework.

Overall, these two books provide a consistent and comprehensive coverage of all aspects of OpRisk management and related insurance analytics as they relate to loss distribution approach modeling and OpRisk – organizational structure, methodologies, policies and infrastructure – for both financial and non-financial institutions. The risk measurement and modeling techniques discussed in the book are based on the latest research. They are presented, however, with considerations based on practical experience of the authors with the daily application of risk measurement tools. We have incorporated the latest evolution of the regulatory framework. The books offer a unique presentation of the latest OpRisk management techniques and provide a unique source of knowledge in risk management ranging from current regulatory issues, data collection and management, technological infrastructure, hedging techniques and organizational structure.

We would like to thank our families for their patience with our absence whilst we were writing this book.

Gareth W. Peters and Pavel V. Shevchenko
London, Sydney, March 2015

Acknowledgments

Dr. Gareth W. Peters acknowledges the support of the Institute of Statistical Mathematics, Tokyo, Japan and Prof. Tomoko Matsui for extended collaborative research visits and discussions during the development of this book.

Acronyms

ABC	approximate Bayesian computation
ALP	accumulated loss policy
a.s.	almost surely
AMA	advanced measurement approach
APT	arbitrage pricing theory
BCRLB	Bayesian Cramer–Rao lower bound
BCBS	Basel Committee on Banking Supervision
BIS	Bank for International Settlements
CV	co-variation
CD	co-difference
CRLB	Cramer–Rao lower bound
CLP	combined loss policy
CVaR	conditional value at risk
DFT	discrete Fourier transform
EVT	extreme value theory
EVI	extreme value index
ES	expected shortfall
FFT	fast Fourier transform
GLM	generalized linear models
GAM	generalized additive models
GLMM	generalized linear mixed models
GAMM	generalized additive mixed models
GAMLSS	generalized additive models for location scale and shape
HMCR	higher moment coherent risk measure
HILP	haircut individual loss policy
ILPU	individual loss policy uncapped
ILPC	individual loss policy capped
i.i.d.	independent and identically distributed
LDA	loss distribution approach
MCMC	Markov chain Monte Carlo
MC	Monte Carlo
MLE	maximum likelihood estimator
MPT	modern portfolio theory
OpRisk	operational risk
PMCMC	particle Markov chain Monte Carlo
r.v.	random variable
SMC	sequential Monte Carlo

SRM	spectral risk measure
SLA	Single Loss Approximation
s.t.	such that
TCE	tail conditional expectation
TTCE	tempered tail conditional expectation
VaR	value at risk
Vco	variational coefficient
w.r.t.	with respect to

Symbols

$\forall$	for all
$\exists$	there exists
$\cup$	union of two sets
$\cap$	intersection of two sets
$*$	convolution operator
$F(\cdot)$	probability distribution function
$C(u_1, u_2, \ldots, u_d)$	d-dimensional Copula probability distribution function.
$\overline{F}$	tail function $\overline{F}(x) = 1 - F(x)$
$f(\cdot)$	probability density function
$h(\cdot)$	hazard rate given by $h(\cdot) = \frac{f(\cdot)}{\overline{F}(\cdot)}$
$\Phi_X[\theta]$	characteristic function for random variable X
$M_X(t)$	moment-generating function of random variable X
$F^{(n)*}(\cdot)$	n-fold convolution of a distribution function with itself
$F^{-1}(\cdot)$	inverse distribution function (quantile function)
$Q(\alpha)$	quantile function
$U(y) = Q(1 - 1/y)$	tail quantile function
$F^{\leftarrow}(\cdot)$	generalized inverse
$g(\cdot) \sim f(\cdot)$	function g is asymptotic equivalent to f at infinity (unless specified otherwise)
$X \sim F(\cdot)$	random variable X is distributed according to F
$X_{(k,n)}$	kth largest sample from n samples, that is, the kth order statistic
$\mathrm{VaR}_\alpha[\cdot]$	value at risk for level α
$\mathrm{ES}_\alpha[\cdot]$	expected shortfall for level α
$\mathrm{SRM}_\alpha[\cdot]$	spectral risk measure for level α
$\mathbb{J}$	space of integers
$\mathbb{R}$	real line
$\mathbb{C}$	complex plane
$\mathscr{R}\mathrm{e}\{\cdot\}$	real component of a complex number or function
$\mathscr{I}\mathrm{m}\{\cdot\}$	imaginary component of a complex number
$\mathbb{E}[\cdot]$	expectation operator
$\mathbb{P}\mathrm{r}[\cdot]$	probability
$f(\cdot) \in RV(\rho)$	function f is in the class of Regularly Varying functions with index of regular variation ρ
$L(\cdot)$	slowly varying function, that is, function $L(\cdot) \in RV(0)$
$\Phi^{(d)}$	dth derivative of function Φ
$\mathbb{I}[\cdot]$	indicator function of an event

$\delta(\cdot)$	Dirac delta function
$o(\cdot)$	little 'Oh' Landau asymptotic notation
$O(\cdot)$	big 'Oh' Landau asymptotic notation
$<<$	Vinogradov's asymptotic notation for big 'Oh' notation
$f \circ g$	f is a composite function of g, that is, $f(g(\cdot))$
$\mathscr{W}[x]$	Lambert W function
$\mathscr{E}[n, x]$	Misra function

List of Distributions

Distribution Name	Distribution Symbol
α-Stable	$S_{\alpha}(\cdot)$
Asymmetric Laplace	$AssymetricLaplace(\cdot)$
Beta	$Beta(\cdot)$
Chi Squared	$ChiSquared(\cdot)$
Dagum or Burr Type XII	$BurrXII(\cdot)$
Double h–h distributions	$T_{h,h}(\cdot)$
Exponential	$Exp(\cdot)$
Gamma	$Gamma(\cdot)$
g distributions	$T_{g}(\cdot)$
g-and-h distributions	$T_{g,h}(\cdot)$
g-and-k distributions	$T_{g,k}(\cdot)$
Generalized gamma	$GG(\cdot)$
Generalized beta	$GB2(\cdot)$
Generalized Hyperbolic	$GH(\cdot)$
Generalized Inverse Guassian	$GIG(\cdot)$
Generalized Log-Logistic	$Gen(\cdot)$
Geometric stable	$GeoStable(\cdot)$
h distributions	$T_{h}(\cdot)$
Halphen type A	$Halphen(\cdot)$
Inverse Gamma	$InverseGamma(\cdot)$
Inverse Gaussian	$InverseGaussian(\cdot)$
Laplace	$Laplace(\cdot)$
Log-Cauchy	$LogCauchy(\cdot)$
LogNormal	$LogNormal(\cdot)$
Log-t	Log-$t(\cdot)$
Lomax	$Lomax(\cdot)$
Normal (Gaussian)	$Normal(\cdot)$
Normal Inverse Gaussian	$NIG(\cdot)$
Singh–Maddala or Burr Type III	$BurrIII(\cdot)$
Standard Normal	$\Phi(\cdot)$
Tempered stable	$TS(\cdot)$

CHAPTER ONE

Motivation for Heavy-Tailed Models

1.1 Structure of the Book

This book is split into a few core components covering fundamental concepts:

- Chapter 1 motivates the need to consider heavy-tailed loss process models in operational risk (OpRisk) and insurance modeling frameworks. It provides a basic introduction to the concept of separating the modeling of the central loss process and the tails of the loss process through splice models. It also sets out the key statistical questions one must consider studying when performing analysis and modeling of high consequence rare-event loss processes.
- Chapter 2 covers all the fundamental properties one may require in univariate loss process modeling under an extreme value theory (EVT) approach. Of importance is the detailed discussion on the associated statistical assumptions that must be made regarding the properties of any data utilized in model estimation when working with EVT models. This chapter provides a relatively advanced coverage of generalized extreme value (GEV) family of models, block maximum and peaks over threshold frameworks. It provides detailed discussion on statistical estimation that should be utilized in practice for such models and how one may adapt such methods to small sample settings that may arise in OpRisk settings. In the process, the chapter details clearly how to construct several loss distributional approach models based on EVT analysis. It then concludes with results of EVT in the context of compound processes.
- Chapter 3 provides a set of formal mathematical definitions for different notions regarding a heavy-tailed or fat-tailed loss distribution and its properties. It is important that when modeling such loss processes, especially the asymptotic properties of compound process

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

models built with heavy-tailed loss models, a clear understanding of the tail properties of such loss models is understood. In this regard, we discuss the family of sub-exponential loss models, the family of regularly varying and slowly varying models. There are within these large classes of models sub-categorizations that are often of use to understand when thinking about risk measures resulting from such loss models, these are also detailed, for example, long-tailed models, subversively varying models and extended regular variation. In addition, the chapter opens with a basic introduction to key notations and properties of asymptotic notations that are utilized throughout the book.

- Chapter 4 begins with a basic introduction to properties of mathematical representations and characterizations of heavy-tailed loss models through the characteristic function and its representation. It then details the notions of divisibility, self-decomposability and the resulting consequences such distributional properties have on loss distributional approach compound process models. The remainder of the chapter provides a detailed coverage of the family of univariate α-stable models, detailing their characterization, the parameterizations, density and distribution representations and parameter estimation. Such a family of models is becoming increasingly interesting for OpRisk modeling and insurance. It is recognized that such a family of models possesses many relevant and useful features that will capture aspects of OpRisk and insurance loss processes accurately and with advantageous features when used in a compound process model under a loss distributional approach structure.
- Chapter 5 provides the representations of flexible severity models based on tempering or exponential tilting of the α-stable family of loss models. Under this concept, there are many families of tempered stable models available; this chapter characterizes each and discusses the mathematical properties of each sub-class of models and how they may be used in compound process models for heavy-tailed loss models in OpRisk and insurance. In addition, it discusses the aspects of model estimation and simulation for such models. The chapter then finishes with a detailed discussion on quantile-transformed-based heavy-tailed loss models for OpRisk and insurance, such as the Tukey transforms and the sub-family of the g-and-h distributions that have been popular in OpRisk.
- Chapter 6 discusses compound processes and convolutional semi-group structures. This then leads to developing representations of closed-form compound process loss distributions and densities that admit heavy-tailed loss processes. The chapter characterizes several classes of such models that can be used in practice, which avoid the need for computationally costly Monte Carlo simulation when working with such models.
- Chapter 7 discusses many properties of different classes of heavy-tailed loss processes with regard to asymptotic representations and properties of the tail of both partial sums and compound random sums. It does so under first-, second- and third-order asymptotic expansions for the tail process of such heavy-tailed loss processes. This is achieved under many different assumptions relating to the frequency and severity distribution and the possible dependence structures in such loss processes.
- Chapter 8 extends the results of the asymptotics for the tail of heavy-tailed loss processes partial sums and compound random sums to the asymptotics of risk measures developed from such loss processes. In particular, it discusses closed-form single-loss approximations and first-order, second-order and higher order expansion representations. It covers value-at-risk, expected shortfall and spectral risk measure asymptotics. This chapter also covers some alternative risk measure asymptotic results based on EVT known as *penultimate approximations*.

- Chapter 9 rounds off the book with a coverage of numerical simulation and estimation procedures for rare-event simulations in heavy-tailed loss processes, primarily for the estimation of properties of risk measures that provides an efficient numerical alternative procedure to utilization of such asymptotic closed-form representations.

1.2 Dominance of the Heaviest Tail Risks

In this book, we develop and discuss models for OpRisk to better understand statistical properties and capital frameworks which incorporate risk classes in which infrequent, though catastrophic or high consequence loss events may occur. This is particularly relevant in OpRisk as can be illustrated by the historical events which demonstrate just how significant the appropriate modeling of OpRisk can be to a financial institution.

Examples of large recent OpRisk losses are

- J.P. Morgan, GBP 3760 million in 2013—US authorities demand money because of mis-sold securities to Fannie Mae and Freddie Mac;
- Madoff and investors, GBP 40,819 million in 2008—B. Madoff's Ponzi scheme;
- Société Générale, GBP 4548 million in 2008—a trader entered futures positions circumventing internal regulations.

Other well-known examples of OpRisk-related events include the 1995 Barings Bank loss of around GBP 1.3 billion; the 2001 Enron loss of around USD 2.2 billion and the 2004 National Australia Bank loss of AUD 360 m.

The impact that such significant losses have had on the financial industry and its perceived stability combined with the Basel II regulatory requirements BCBS (2006) have significantly changed the view that financial institutions have regarding OpRisk. Under the three pillars of the Basel II framework, internationally active banks are required to set aside capital reserves against risk, to implement risk management frameworks and processes for their continual review and to adhere to certain disclosure requirements. There are three broad approaches that a bank may use to calculate its minimal capital reserve, as specified in the first Basel II pillar. They are known as *basic indicator approach*, *standardized approach* and *advanced measurement approach* (AMA) discussed in detail in Cruz *et al.* (2015). AMA is of interest here because it is the most advanced framework with regards to statistical modeling. A bank adopting the AMA must develop a comprehensive internal risk quantification system. This approach is the most flexible from a quantitative perspective, as banks may use a variety of methods and models, which they believe are most suitable for their operating environment and culture, provided they can convince the local regulator (BCBS 2006, pp. 150–152). The key quantitative criterion is that a bank's models must sufficiently account for potentially high impact rare events. The most widely used approach for AMA is loss distribution approach (LDA) that involves modeling the severity and frequency distributions over a predetermined time horizon so that the overall loss Z of a risk over this time period (e.g. year) is

$$Z = X_1 + \cdots + X_N, \tag{1.1}$$

where N is the frequency modeled by random variable from discrete distribution and $X_1, X_2, \ldots$ the independent severities from continuous distribution $F_X(x)$. There are many important aspects of LDA such as estimation of frequency and severity distributions using

data and expert judgements or modeling dependence between risks considered in detail in Cruz *et al.* (2015). In this book, we focus on modeling heavy-tailed severities.

Whilst many OpRisk events occur frequently and with low impact (indeed, are 'expected losses'), others are rare and their impact may be as extreme as the total collapse of the bank. The modeling and development of methodology to capture, classify and understand properties of operational losses is a new research area in the banking and finance sector. These rare losses are often referred to as *low frequency/high severity risks*. It is recognized that these risks have heavy-tailed (sub-exponential) severity distributions, that is, the distribution with the tail decaying to zero slower than any exponential.

In practice, heavy-tailed loss distribution typically means that the observed losses are ranging over several orders of magnitude, even for relatively small datasets. One of the main properties of heavy-tailed distributions is that if $X_1, \ldots, X_n$ are independent random variables from common heavy-tailed distribution $F(x)$, then

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x]}{\mathbb{P}\mathrm{r}[\max(X_1, \ldots, X_n) > x]} = 1. \tag{1.2}$$

This means that the tail of the sum of the random variables has the same order of magnitude as the tail of the maximum of these random variables, with interpretation that severe overall loss is due to a single large loss rather than due to accumulated small losses.

In OpRisk and insurance, we are often interested in the tail of distribution for the overall loss over a predetermined time horizon $Z = X_1 + \cdots + X_N$. In this case, if $X_1, X_2, \ldots$ are independent severities from heavy-tailed distribution $F_X(x)$ and

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n \mathbb{P}\mathrm{r}[N = n] < \infty$$

for some $\varepsilon > 0$ (which is satisfied, e.g. for Poisson and negative binomial distributions), then

$$1 - F_Z(x) = \mathbb{P}\mathrm{r}[X_1 + \cdots + X_N > x] \sim \mathbb{E}[N](1 - F_X(x)), \quad x \to \infty. \tag{1.3}$$

This can be used to approximate high quantiles of the distribution of Z as

$$F_Z^{-1}(q) \sim F_X^{-1}\left(1 - \frac{1-q}{\mathbb{E}[N]}\right), \quad q \to 1, \tag{1.4}$$

where q is the quantile level. This approximation is often referred to as the *single-loss approximation* because the compound distribution is expressed in terms of the single-loss distribution.

Heavy-tailed distributions include many well-known distributions. For example, the Log-Normal distribution is heavy tailed. An important class of heavy-tailed distributions is the so-called regular varying tail distributions (often referred to as *power laws* or *Pareto distributions*)

$$1 - F(x) = x^{-\alpha} C(x), \quad x \to \infty, \ \alpha \geq 0, \tag{1.5}$$

where α is the so-called power tail index and $C(x)$ the slowly varying function that satisfies

$$\lim_{x\to\infty} C(tx)/C(x) = 1, \text{ for } t > 0. \tag{1.6}$$

Often, sub-exponential distributions provide a good fit to the real datasets of OpRisk and insurance losses. However, corresponding datasets are typically small and the estimation of these distributions is a difficult task with a large uncertainly in the estimates.

Remark 1.1 *From the perspective of capital calculation, the most important processes to model accurately are those which have relatively infrequent losses. However, when these losses do occur, they are distributed as a very heavy-tailed severity distribution such as members of the sub-exponential family. Therefore, the intention of this book is to present families of models suitable for such severity distribution modeling as well as their properties and estimators for the parameters that specify these models.*

The precise definition and properties of the heavy-tailed distributions is a subject of Chapter 3, and single-loss approximation is discussed in detail in Chapters 7 and 8. For a methodological insight, consider J independent risks, where each risk is modeled by a compound Poisson. Then, the sum of risks is a compound Poisson with the intensity and severity distribution given by the following proposition.

Proposition 1.1 *Consider J independent compound Poisson random variables*

$$Z^{(j)} = \sum_{s=1}^{N^{(j)}} X_s^{(j)}, \quad j = 1, \ldots, J, \tag{1.7}$$

where the frequencies $N^{(j)} \sim Poisson(\lambda_j)$ and the severities $X_s^{(j)} \sim F_j(x)$, $j = 1, \ldots, J$ and $s = 1, 2, \ldots$ are all independent. Then, the sum $Z = \sum_{j=1}^{J} Z^{(j)}$ is a compound Poisson random variable with the frequency distribution $Poisson(\lambda)$ and the severity distribution

$$F(x) = \sum_{j=1}^{J} \frac{\lambda_j}{\lambda} F_j(x),$$

where $\lambda = \lambda_1 + \cdots + \lambda_J$.

The proof is simple and can be found, for example, in Shevchenko (2011, proposition 7.1). Suppose that all severity distributions $F_j(x)$ are heavy tailed, that is,

$$\overline{F}_j(x) = x^{-\alpha_j} C_j(x),$$

where $\alpha_1 < \cdots < \alpha_J$ and $C_j(x)$ are the slowly varying functions as defined in Equation 1.6. Then, $F(x) = \sum_{j=1}^{J} (\lambda_j/\lambda) F_j(x)$ is a heavy-tailed distribution too, with the tail index α_1 for $x \to \infty$. Thus, using the result (Equation 1.3) for heavy-tailed distributions, we obtain that

$$\lim_{x \to \infty} \frac{\mathbb{P}\mathrm{r}[Z > x]}{1 - F_1(x)} = \lambda_1. \tag{1.8}$$

This means that high quantiles of the total loss are due to the high losses of the risk with the heaviest tail. For illustration of this phenomenon with the real data from ORX database, see Cope *et al.* (2009). In their example, $LogNormal\ (\mu = 8, \sigma = 2.24)$ gave a good fit for 10 business lines with average 100 losses per year in each line using 10,000 observations. The estimated capital across these 10 business lines was Euro 634 million with 95% confidence

interval (uncertainty in the capital estimate due to finite data size) of width Euro 98 million. Then, extra risk cell (corresponding to the "clients, products & business practices" event type in the 'corporate finance' business line) was added with one loss per year on an average and the $LogNormal(\mu = 9.67, \sigma = 3.83)$ severity estimated using 300 data points. The obtained estimate for the capital over the 10 business units plus the additional one was Euro 5260 million with 95% confidence interval of the width Euro 19 billion. This shows that one high severity risk cell contributes 88% to the capital estimate and 99.5% to the uncertainty range. In this example, the high severity unit accounts for 0.1% of the bank's losses.

Another important topic in modeling large losses is EVT that allows to extrapolate to losses beyond those historically observed and estimate their probability. There are two types of EVT: *block maxima* and *threshold exceedances*; both are considered in Chapter 2. EVT block maxima are focused on modeling the largest loss per time period of interest. Modeling of all large losses exceeding a large threshold is dealt by EVT threshold exceedances. The key result of EVT is that the largest losses or losses exceeding a large threshold can be approximated by some limiting distributions which are the same regardless of the underlying process. This allows to extrapolate to losses beyond those historically observed. However, EVT is an asymptotic theory. Whether the conditions validating the use of the asymptotic theory are satisfied is often a difficult question to answer. The convergence of some parametric models to EVT regime is very slow. In general, it should not preclude the use of other parametric distributions. In Chapter 4, we consider many useful flexible parametric heavy-tailed distributions.

It is important to mention that empirical data analysis for OpRisk often indicates stability of an infinite mean model for some risk cells (e.g. see Moscadelli (2004)), that is, the severity distribution is a Pareto-type distribution (Equation 1.5) with $0 < \alpha \leq 1$ that has infinite mean. For a discussion about infinite mean models in OpRisk, see discussions in Nešlehová *et al.* (2006). Often, practitioners question this type of model and apply different techniques such as truncation from the above but then the high quantiles become highly dependent on the cut-off level. Typically, the estimates of high quantiles for fat-tailed risks have a very large uncertainty and the overall analysis is less conclusive than in the case of thin-tailed risks; however, it is not the reason to avoid these models if the data analysis points to heavy-tailed behaviour. Recent experience of large losses in OpRisk, when one large loss may lead to the bankruptcy, certainly highlights the importance of the fat-tailed models.

1.3 Empirical Analysis Justifying Heavy-Tailed Loss Models in OpRisk

There are several well-known published empirical studies of OpRisk data such as Moscadelli (2004) analysing 2002 Loss Data Collection Exercise (LDCE) survey data across 89 banks from 19 countries; Dutta & Perry (2006) analysing 2004 LDCE for US banks and Lu & Guo (2013) analysing data in Chinese banks.

- Moscadelli (2004) analysed 2002 Loss Data Collection Exercise (LDCE) survey data with more than 47,000 observations across 89 banks from 19 countries in Europe, North and South Americas, Asia and Australasia. The data were mapped to the Basel II standard eight business lines and seven event types. To model severity distribution, this study considered generalized Pareto distribution (EVT distribution for threshold exceedances in the limit of large threshold) and many standard two-parameter distributions such as gamma,

exponential, Gumbel and LogNormal. The analysis showed that EVT explains the tail behaviour of OpRisk data well.

- Dutta & Perry's (2006) study of US banking institutions considered the 2004 LDCE survey data and narrowed down the number of suitable candidate datasets from all institutions surveyed to just seven institutions for which it was deemed sufficient numbers of reported losses were acquired. The somewhat heuristic selection criterion that the authors utilized was that a total of at least 1,000 reported total losses were required and, in addition, each institution was required to have consistent and coherent risk profiles relative to each other, which would cover a range of business types and risk types as well as asset sizes for the institutions.
- Feng *et al.*'s (2012) study on the Chinese banking sector utilized less reliable data sources for loss data of Chinese commercial banks collected through the national media covering 1990–2010. In the process collecting data for banks which include the 4 major state-owned commercial banks (SOCBs), 9 joint-stock commercial banks (JSCBs), 35 city commercial banks (CCBs), 74 urban and rural credit cooperatives (URCCs) and 13 China Postal Savings subsidiaries (CPS). The authors also note that the highest single OpRisk loss amount is up to 7.4 billion yuan, whereas the lowest amount is 50,000 yuan. In addition, losses measured in foreign currency were converted back via the real exchange rate when the loss occurred to convert it to the equivalent amount in yuan. Details of the incidence bank, incidence bank location, type of OpRisk loss, amount of loss, incident time and time span and the sources of OpRisk events were noted.

In the following, we focus on the study of Dutta & Perry (2006), where the authors explored a number of key statistical questions relating to the modeling of OpRisk data in practical banking settings. As noted, a key concern for banks and financial institutions, when designing an LDA model, is the choice of model to use for modeling the severity (dollar value) of operational losses. In addition, a key concern for regulatory authorities is the question of whether institutions using different severity modeling techniques can arrive at very different (and inconsistent) estimates of their exposure. They found, not surprisingly, that using different models for the same institution can result in materially different capital estimates. However, on the more promising side for LDA modeling in OpRisk, they found that there are some models that yield consistent and plausible results for different institutions even when their data differs in some core characteristics related to the collection processes. This suggests that OpRisk data displays some regularity across institutions which can be modeled. In this analysis, the authors noted that they were careful to consider both the modeling of aggregate data at the enterprise level, which would group losses from different business lines and risk types and modeling the attributes of the individual business line and risk types under the recommended business lines of Basel II/Basel III.

On the basis of data collected from seven institutions, with each institution selected as it had at least 1,000 loss events in total, and the data was part of the 2004 LDCE, they performed a detailed statistical study of attributes of the data and flexible distributional models that could be considered for OpRisk models. On the basis of these seven data sources, over a range of different business units and risk types, they found that fitting all of the various datasets one would need to use a model that is flexible enough in its structure. Dutta & Perry (2006) considered modeling via several different means: parametric distributions, EVT models and non-parametric empirical models.

The study focused on models considered by financial institutions in Quantitative Impact Study 4 (QIS-4) submissions, which included one-, two- and four-parameter models. The

one- and two-parameter distributions for the severity models included exponential, gamma, generalized Pareto, LogLogistic, truncated LogNormal and Weibull. The four-parameter distributions include the generalized Beta distribution of second kind (GB2) and the g-and-h distribution. These models were also considered in Peters & Sisson (2006a) for modeling severity models in OpRisk under a Bayesian framework.

Dutta & Perry (2006) discussed the importance of fitting distributions that are flexible but appropriate for the accurate modeling of OpRisk data; they focussed on the following five simple attributes in deciding on a suitable statistical model for the severity distribution.

1. *Good Fit*. Statistically, how well does the model fit the data?
2. *Realistic*. If a model fits well in a statistical sense, does it generate a loss distribution with a realistic capital estimate?
3. *Well Specified*. Are the characteristics of the fitted data similar to the loss data and logically consistent?
4. *Flexible*. How well is the model able to reasonably accommodate a wide variety of empirical loss data?
5. *Simple*. Is the model easy to apply in practice, and is it easy to generate random numbers for the purposes of loss simulation?

Their criterion was to regard any technique that is rejected as a poor statistical fit for the majority of institutions to be inferior for modeling OpRisk. The reason for this consideration was related to their desire to investigate the ability to find aspects of uniformity or universality in the OpRisk loss process that they studied. From the analysis undertaken, they concluded that such an approach would suggest OpRisk can be modeled, and there is regularity in the loss data across institutions. Whilst this approach combined elements of expert judgement and statistical hypothesis testing, it was partially heuristic and not the most formal statistical approach to address such problems. However, it does represent a plausible attempt given the limited data sources and resources as well as competing constraints mentioned in the measurement criterion they considered.

We note that an alternative purely statistical approach to such model selection processes was proposed in OpRisk modeling in the work of Peters & Sisson (2006a), whose approach to model selection was to consider a Bayesian model selection based on Bayesian methodology of the Bayes factor and information criterion for penalized model selection such as the Bayesian information criterion.

In either approach, it is generally acknowledged that accurate model selection of an appropriate severity model is paramount to appropriate modeling of the loss processes and, therefore, to the accurate estimation of capital.

Returning to the findings from the seven sources of OpRisk data studied in Dutta & Perry (2006), they found that the exponential, gamma and Weibull distributions are rejected as good fits to the loss data for virtually all institutions at the enterprise, business line and event type levels. This was decided based on formal one sample statistical goodness of fit tests for these models.

When considering the g-and-h distribution, they did not perform the standard hypothesis test for goodness of fit instead opting for a comparison of quantile–quantile (Q–Q) plots and diagnostics based on the five criteria posed above. In all the situations, they found that the g-and-h distribution fits as well as other distributions on the Q–Q plot. The next most preferred distributions were the GB2, LogLogistic, truncated LogNormal and generalized Pareto

models, indicating the importance of considering flexible severity loss models. However, only g-and-h distribution resulted in realistic and consistent capital estimates across all seven institutions. In addition, they noted that the EVT models fitted under an EVT threshold exceedances framework were also generally suitable fits for the tails, consistent with the discussions and findings in Lu & Guo (2013) for OpRisk data in the Chinese banking sector and with the results in Moscadelli (2004) analysing 2002 LDCE.

1.4 Motivating Parametric, Spliced and Non-Parametric Severity Models

In this section, we discuss the different approaches that have been adopted in the literature to model aspects of heavy-tailed loss processes. Primarily we focus on the modeling of the severity process in an OpRisk LDA framework; however, we note that many of these approaches can also be adopted for modeling of the annual loss process should sufficient data be available. Before discussing these approaches, it is important to understand some of the basic implications associated with subscribing to such modeling frameworks. We detail two of the most fundamental of these in the following.

Basic Statistical Assumptions to be Considered in Practice

1. It is typical from the statistical perspective to apply the models to be discussed later on the proviso that the underlying process under consideration is actually arising from a single physical process responsible for the losses to be observed. However, in practice, several authors have discussed the impracticality of such assumptions in real-world financial environments, which unlike their physical extreme process counterparts often studied in atmospheric science, hydrology and meteorology, such financial processes are difficult to attribute to a fundamental single 'physical' driving force. Discussion on such issues and their consequences to the suitability of such modeling approaches is provided in Cope *et al.* (2009) and Chavez-Demoulin *et al.* (2006).
2. The other typical statistical assumption that will have potential consequences to application of such modeling paradigms to be discussed later relates to the assumptions made on the temporal characteristics of the underlying loss process driving the heavy-tailed behaviour. In most modeling frameworks discussed later, the parameters causing the loss process will typically be considered unknown but static over time. However, it is likely that in dynamically evolving commercial environments in which financial institutions, disappear, appear and merge on a global scale, whilst regulation continually adapts to the corporate and political landscape, such loss processes driving the heavy-tailed behaviour may not have parameters which are static over time. For example, it is common that under severe losses from an event such as rogue trading, one would typically see the financial institution involved take significant measures to modify the process with the aim to prevent such losses in the same manner again in future, by changing the financial controls, policies and regulatory oversight. This has practical consequences for the ability to satisfy the typical statistical assumptions one would like to adopt with such heavy-tailed models.
3. Typically, the application and development of theoretical properties of the models to be developed, including the classical estimators developed for the parameters of such models under either a frequentist or a Bayesian modeling paradigm, revolve around the assumption that the losses observed are independent and identically distributed. Again,

several authors have developed frameworks motivating the necessity to capture dependence features adequately in OpRisk and insurance modeling of heavy-tailed data, see Böcker & Klüppelberg (2008), Chavez-Demoulin *et al.* (2006) and Peters *et al.* (2009a). In practice, the models presented later can be adapted to incorporate dependence, once a fundamental understanding of their properties and representations is understood for the independently and identically distributed (i.i.d.) cases and this is an active field of research in OpRisk at present.

4. Finally, there is also, typically for several high consequence loss processes, a potential upper limit of the total loss that may be experienced by such a loss process. Again, this is practically important to consider before developing such models to be presented.

The actuarial literature has undertaken several approaches to attempt to address aspects of modeling when such assumptions are violated. For example, should one believe that the underlying risk process is a consequence of multiple driving exposures and processes, then it is common to develop what are known as mixture loss processes. Where if one can identify key loss processes that are combining to create the observed loss process in the OpRisk framework under study, then fitting a mixture model in which there is one component per driving process (potentially with different heavy-tailed features) is a possibility. Another approach that can be adopted and we discuss in some detail throughout next section is the method known as *splicing*. In such a framework, a flexible severity distribution is created, which aims to account for two or more driving processes that give rise to the observed loss process. This is achieved under a splicing framework under the consideration that the loss processes combining to create the observed process actually may differ significantly in the amounts of losses they generate and also in OpRisk perhaps in the frequency at which these losses are observed. Therefore, a splicing approach adopts different models for particular intervals of the observed loss magnitudes. Therefore, small losses may be modeled by one parametric model over a particular interval of loss magnitudes and large severe losses captured by a second model fitted directly to the losses observed in the adjacent loss magnitude partition of the loss domain. These will be discussed in some detail in the following chapter.

In general, it is a serious challenge for the risk managers in practice to try to reconcile such assumptions into a consistent, robust and defensible modeling framework. Therefore, we proceed with an understanding that such assumptions may not all be satisfied jointly under any given model when developing the frameworks to be discussed later. However, in several cases, the models we will present will in many respects provide a conservative modeling framework for OpRisk regulatory reporting and capital estimation should these assumptions be violated as discussed earlier.

Statistical Modeling Approaches to Heavy-Tailed Loss Processes:

The five basic statistical approaches to modeling the severity distribution for a single-loss process that will be considered throughout this book are:

1. EVT methods for modeling explicitly the tail behaviour of the severity distribution in the loss process: 'block maxima' and 'points over threshold' models.
2. Spliced parametric distributional models combining exponential family members with EVT model tail representations: mixtures and composite distributions.
3. Spliced non-parametric kernel density estimators with EVT tail representations.
4. Flexible parametric models for the entire severity distribution considered from sub-exponential family members: α-stable, tempered and generalized tempered α-stable,

generalized hypergeometric (normal inverse Gaussian), GB2, generalized Champernowne and quantile distributions (g-and-h).

5. Spliced parametric distributional models examples combining exponential family members with sub-exponential family parametric models.

As is evident from the survey of different approaches to modeling heavy-tailed loss processes, mentioned earlier, there is a large variety of models and techniques developed to study and understand such important phenomena as heavy-tailed processes. In the context of OpRisk, the consequences of failing to model adequately the possible heavy-tailed behaviour of certain OpRisk loss processes could result in significant under estimation of the required capital to guard against realizations of such losses in a commercial banking environment and the subsequent failure or insolvency of the institution.

1.5 Creating Flexible Heavy-Tailed Models via Splicing

In this section, we briefly detail the basic approaches to create a spliced distribution and the motivation for such models. These will then be significantly elaborated in the proceeding models when they are incorporated with various modeling approaches to capture heavy-tailed behaviour of a loss process.

It is common in practice for actuarial scientist and risk managers to consider the class of flexible distributional models known as *spliced distributions*. In fact, there are standard packages implemented in several widely utilized software platforms for statistical and risk modeling that incorporate at least basic features of spliced models. The basic k-component spliced distribution as presented in Klugman *et al.* (1998, section 5.2.6) is defined according to Definition 1.1.

Definition 1.1 (Spliced Distribution) *A random variable $X \in \mathbb{R}^+$ representing the loss of a particular risk process can be modeled by a k-component spliced distribution, defined according to the density function partitioned over the loss magnitudes according to the intervals $\cup_{i=1}^{k}[x_{i-1}, x_i) = \mathbb{R}^+$ and given by*

$$f_X(x) = \begin{cases} w_1\, f_1(x), & 0 \le x < x_1, \\ w_2\, f_2(x), & x_1 \le x < x_2, \\ \vdots & \\ w_{k-1}\, f_{k-1}(x), & x_{k-2} \le x < x_{k-1}, \\ w_k\, f_k(x), & x_{k-1} \le x < \infty, \end{cases} \tag{1.9}$$

where the weight parameters $w_i \ge 0$, $i = 1, \ldots, k$ satisfy $w_1 + \cdots + w_k = 1$, and $f_1(x), \ldots, f_k(x)$ are proper density functions, that is, $\int f_i(x)dx = 1$, $i = 1, \ldots, k$.

To illustrate this, consider the typically applied model involving the choice of $k = 2$ in which the loss processes have loss magnitudes which are partitioned into two regions $[0, x_{\min}) \cup [x_{\min}, \infty)$. The interpretation being that two driving processes give rise to the risk processes under study. Less frequent but more severe loss processes would typically experience

losses exceeding $x_{\min}$. Therefore, we may utilise a lighter tailed parametric model $f_1(x)$ in the region $[0, x_{\min})$ and an associated normalization for the truncation of the distribution over this region. This would be followed by a heavier tailed perhaps parametric model $f_2(x)$ in the region $[x_{\min}, \infty)$, which would also be standardized by w_2 to ensure that the total resulting density on $\mathbb{R}^+$ was appropriately normalized. Clearly, there are several approaches that can be adopted to achieve this, for example, one may wish to ensure continuity or smoothness of the joint distribution such as at the boundary points between adjacent partitions. This will impose restrictions on the parameters controlling the distributional models; in other settings, such concerns will not be of consequence. Example illustrations of such models are provided in Examples 1.1–1.4, which illustrate a discontinuous model and continuous models, respectively.

EXAMPLE 1.1 Parametric Body and Parametric Tail

Assume that losses $X_1, X_2, \ldots, X_K$ are independent and identically distributed. If we want to model the losses above a selected threshold $x_{\min}$ using some parametric distribution $G_2(x)$ with density $g_2(x)$ defined on $x > 0$ (e.g. LogNormal distribution) and the losses below using another parametric distribution $G_1(x)$ with density $g_1(x)$ defined on $x > 0$ (e.g. Gamma distribution), then corresponding density $f(x)$ and distribution $F(x)$ for spliced model to fit are

$$f(x) = wf_1(x) + (1-w)f_2(x),$$

$$F(x) = \begin{cases} wF_1(x), & 0 < x < x_{\min}, \\ w + (1-w)F_2(x), & x \geq x_{\min}, \end{cases}$$

where $w \in [0,1]$ is the weight parameter and the proper densities $f_1(x)$ and $f_2(x)$ (and their distribution functions $F_1(x)$ and $F_2(x)$) correspond to the densities $g_1(x)$ and $g_2(x)$ truncated above and below $x_{\min}$, respectively:

$$f_1(x) = \frac{g_1(x)}{G_1(x_{\min})}\mathbb{I}_{x<x_{\min}}, \quad F_1(x) = \frac{G_1(x)}{G_1(x_{\min})},\ x < x_{\min},$$

$$f_2(x) = \frac{g_2(x)}{1-G_2(x_{\min})}\mathbb{I}_{x\geq x_{\min}}, \quad F_2(x) = \frac{G_2(x)-G_2(x_{\min})}{1-G_2(x_{\min})},\ x \geq x_{\min}.$$

■

EXAMPLE 1.2 Empirical Body and Parametric Tail

Assume that losses $X_1, X_2, \ldots, X_K$ are independent and identically distributed. If we want to model the losses above a selected threshold $x_{\min}$ using some parametric distribution $G_2(x)$ with density $g_2(x)$ defined on $x > 0$ (e.g. LogNormal distribution) and the losses below using empirical distribution

$$G_1(x) = \frac{1}{K}\sum_{k=1}^{K}\mathbb{I}_{X_k \leq x},$$

then corresponding distribution $F(x)$ for the spliced model is

$$F(x) = \begin{cases} G_1(x), & 0 < x < x_{\min}, \\ G_1(x_{\min}) + (1 - G_1(x_{\min}))F_2(x), & x \geq x_{\min}, \end{cases}$$

where $F_2(x)$ is distribution $G_2(x)$ truncated below $x_{\min}$, that is,

$$F_2(x) = \frac{G_2(x) - G_2(x_{\min})}{1 - G_2(x_{\min})} \mathbb{I}_{x \geq x_{\min}}.$$

Comparing to Example 1.1, note that we selected weight parameter $w = G_1(x_{\min})$ to have a model consistent with the data below $x_{\min}$.

If the threshold $x_{\min}$ is large enough, then (under the regularity conditions of EVT threshold exceedances discussed in Chapter 2 and satisfied for most of the distributions used in practice), the truncated distribution $F_2(x)$ may be approximated by the generalized Pareto distribution

$$G_{\xi,\beta}(x - x_{\min}) = \begin{cases} 1 - (1 + \xi(x - x_{\min})/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-(x - x_{\min})/\beta), & \xi = 0. \end{cases}$$ ■

EXAMPLE 1.3 Gamma Body and Pareto Tail, Discontinuous Density

Consider a loss process with loss random variable $X \sim f_X(x)$ modeled according to a $k = 2$ component spliced distribution comprised of a gamma distribution over the intervals $[0, x_{\min})$ and a Pareto distribution over the interval $[x_{\min}, \infty)$. The resulting density, without any continuity restrictions at the partition boundary, is, therefore, given by

$$f_X(x) = \begin{cases} w_1 Z_1^{-1} \dfrac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), & x \in [0, x_{\min}), \alpha, \beta > 0 \\ w_2 \dfrac{\gamma x_{\min}^{\gamma}}{x^{\gamma+1}}, & x \in [x_{\min}, \infty),\ \gamma > 0 \end{cases} \tag{1.10}$$

with

$$\begin{aligned} Z_1 = F_X(x; \alpha, \beta) &= \int_0^{x_{\min}} \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \\ &= \Gamma^{-1}(\alpha) \underbrace{\int_0^{x/\beta} t^{\alpha-1} \exp(-t)\, dt}_{\text{lower incomplete Gamma function}} \\ &= \Gamma^{-1}(\alpha) \gamma\left(\alpha, \frac{x}{\beta}\right) \end{aligned} \tag{1.11}$$

and subject to the constraint that $\sum_{i=1}^{2} w_i = 1$. Furthermore, we may wish to consider cases typically in practice in which the mode of the first partitions distribution lies in the interval $(0, x_{\min})$ in which cases we further impose the restriction on the shape and scale parameters such that $(\alpha - 1)\beta \in (0, x_{\min})$. In the illustration, we consider the

resulting density for the settings $x_{\min} = 100$, $w_1 = 0.5$, $\alpha = 2, \beta = 10$ and $\gamma = 0.4$ giving a density in Figure 1.1.

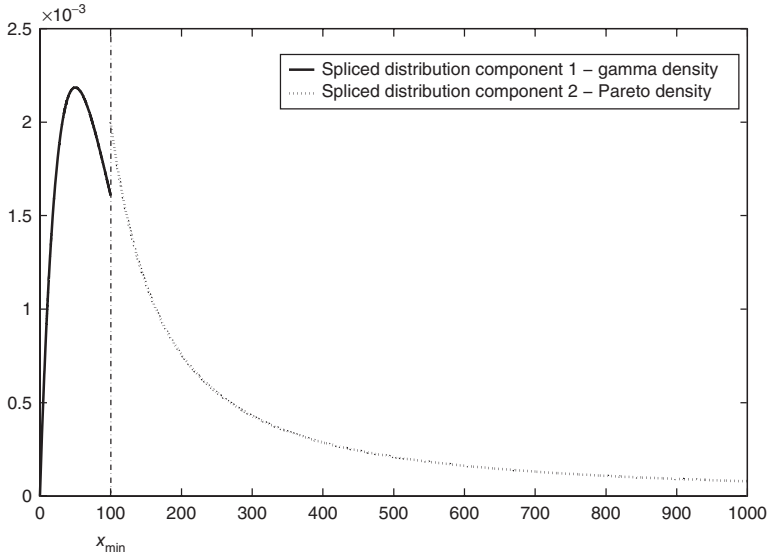

FIGURE 1.1 **Spliced density plot for gamma and Pareto distribution single risk severity model, no continuity constraints.**

EXAMPLE 1.4 Gamma Body and Pareto Tail, Continuous Density

Consider a loss process with loss random variable $X \sim f_X(x)$ modeled according to a $k = 2$ component spliced distribution comprised again of a gamma distribution over the intervals $[0, x_{\min})$ and a Pareto distribution over the interval $[x_{\min}, \infty)$. This time the resulting density is developed subject to the constraint that a certain degree of smoothness is present at the partition boundary connecting the two density functions as captured by equality of the first moments. The resulting density is then developed according to

$$f_X(x) = \begin{cases} w_1 Z_1^{-1} \dfrac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\dfrac{x}{\beta}\right), & x \in [0, x_{\min}), \alpha, \beta > 0, \\ w_2 \dfrac{\gamma x_{\min}^{\gamma}}{x^{\gamma+1}}, & x \in [x_{\min}, \infty), \gamma > 0, \end{cases} \tag{1.12}$$

with

$$\begin{aligned} Z_1 = F_X(x; \alpha, \beta) &= \int_0^{x_{\min}} \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \\ &= \Gamma^{-1}(\alpha) \underbrace{\int_0^{x/\beta} t^{\alpha-1} \exp(-t) dt}_{\text{lower incomplete gamma function}} \\ &= \Gamma^{-1}(\alpha) \gamma \left(\alpha, \frac{x}{\beta}\right) \end{aligned} \tag{1.13}$$

and subject to the constraint that $\sum_{i=1}^{2} w_i = 1$. Furthermore, we may wish to consider cases typically in practice in which the mode of the first partitions distribution lies in the interval $(0, x_{\min})$ in which cases we further impose the restriction on the shape and scale parameters such that $(\alpha - 1)\beta \in (0, x_{\min})$. In addition, the continuity constraint implies the following restriction on the two densities at $x_{\min}$,

$$\begin{aligned} f_1(x_{\min}) &= f_2(x_{\min}) \\ \left.\frac{df_1(x)}{dx}\right|_{x=x_{\min}} &= \left.\frac{df_2(x)}{dx}\right|_{x=x_{\min}}. \end{aligned} \tag{1.14}$$

Where these two restrictions create the following system of constraints that the model parameters must satisfy

$$\begin{aligned} f_1(x_{\min}) - f_2(x_{\min}) &= w_1 Z_1^{-1} \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x_{\min}^{\alpha-1} \exp\left(-\frac{x_{\min}}{\beta}\right) - w_2 \frac{\gamma x_{\min}^{\gamma}}{x_{\min}^{\gamma+1}} = 0 \\ \left.\frac{df_1(x)}{dx}\right|_{x=x_{\min}} - \left.\frac{df_2(x)}{dx}\right|_{x=x_{\min}} &= \frac{w_1}{Z_1} \frac{\beta^{-\alpha}}{\Gamma(\alpha)} \left[(\alpha-1)x_{\min}^{\alpha-2} \exp\left(-\frac{x_{\min}}{\beta}\right)\right. \\ &\left. -\frac{1}{\beta} x_{\min}^{\alpha-1} \exp\left(-\frac{x_{\min}}{\beta}\right)\right] + w_2 \gamma x_{\min}^{-2}(\gamma+1) = 0. \end{aligned} \tag{1.15}$$

In the following example illustration, we consider the resulting density for the settings $x_{\min} = 100$, $w_1 = 0.5$, $\alpha = 2$ and β, γ each set to satisfy these constraints giving a density given in Figure 1.2.

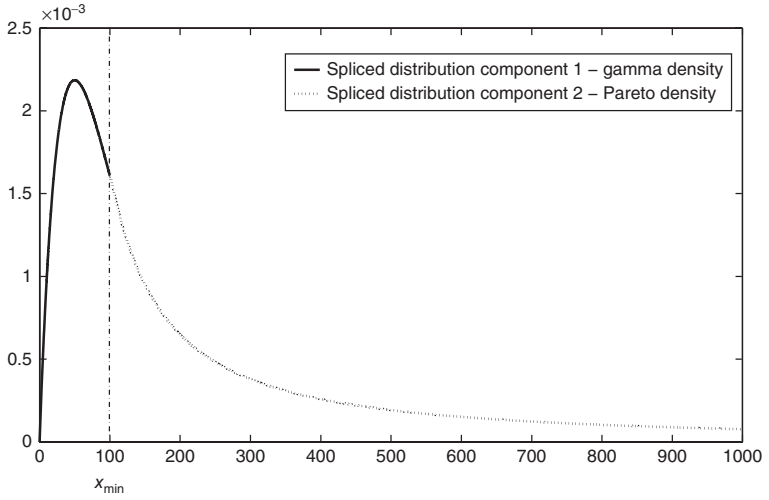

FIGURE 1.2 **Spliced density plot for Gamma and Pareto distribution single risk severity model, with continuity constraints.**

CHAPTER TWO

Fundamentals of Extreme Value Theory for OpRisk

2.1 Introduction

In practical scenarios, it is standard practice to consider basic two-parameter models in the operational risk (OpRisk) modeling of a single-loss process severity distribution under a loss distribution approach (LDA) approach. The most common of these models is based around a LogNormal distribution. Part of the reason for this is the inherent simplicity with which the estimation of the model parameters in this model can be performed. However, because of the extreme quantile requirements that are required to be obtained in regulatory reporting of capital under the Basel II and Basel III accords, the use of models such as LogNormal may not adequately capture the tail features of the loss process under study at such high quantile levels. Therefore, the intention of this section involves motivation of extreme value theory (EVT) concepts to study and understand such extreme loss behaviour.

In this regard, we also note that the concept of what constitutes a heavy-tailed distribution for severity modeling in OpRisk can be defined in several different ways from a statistical perspective. It is common to consider the probabilistic definition of a heavy-tailed distribution as those probability distributions whose tails are not exponentially bounded. That is, they have heavier tails than the exponential distribution. In many applications, it is the right tail of the distribution that is of interest, but a distribution may have a heavy left tail or both tails may be heavy. In the context of OpRisk modeling, such distributions will model explicitly either the severity distribution or the resulting annual loss process distribution for the risk process. As such we will focus naturally on distributions with positive support, and because we are considering loss distributions, the main emphasis will be on the right tail. There are two important subclasses of heavy-tailed distributions, the long-tailed distributions and the

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

sub-exponential distributions. In practice, all commonly used heavy-tailed distributions belong to the sub-exponential class.

We note that there are two common alternative definitions that one may also find in the literature, defining what constitutes a heavy-tailed distribution, these include distributions which do not have all their power moments finite and, in some instances, distributions that do not have a finite variance are also referred to as *heavy tailed*.

We begin this section by first introducing fundamental concepts of EVT which studies the properties and distributional results related to the maximum loss that may occur when the risk process has observed losses that arise from heavy-tailed distributions. The results that arise from this theoretical background are of direct relevance to concepts utilized in the study of properties of the annual loss distribution in OpRisk in the setting of heavy-tailed severity distributions.

Next we present a very general definition of a heavy-tailed class of distribution based on the notion of a sub-exponential family as it includes all distributions encompassed by the alternative definitions based on finite moment properties (e.g. Pareto distribution) as well as those distributions that possess all their power moments (e.g. LogNormal distribution), yet which are generally acknowledged to be heavy tailed.

2.2 Historical Perspective on EVT and Risk

EVT is conceptually an appealing and elegant theoretical study of the heavy-tailed behaviour of a large number of applied physical domains. Traditionally, such applications have involved domains of civil engineering and design of buildings and structures, air pollution and environmental monitoring, hydrology and the design of dams and dyke's. Common to most of these applications is the collection of observed data from such physical processes spatially and/or temporally collected. The resulting modeling of such data then focuses on the extremes observed. In this chapter, we focus our analysis not on such physical phenomena but instead on the analysis and understanding of EVT from the perspective of a risk management application. In such environments, there is less interpretation of the fundamental 'physical' process that gives rise to the extremal values, in this case losses. Nevertheless, there is still a strong motivation in understanding and modeling such processes to consider the well-established theory that EVT provides.

A second challenge faced in the consideration of OpRisk modeling under an EVT framework is that typically the observed extreme losses are few and far between. The data sources can be significantly scarcer than those available in other domains in which EVT has been developed. This poses a significant challenge that is partially addressed by risk managers through either Bayesian modeling approaches or scenario analysis.

Additionally, the other novelty that is present in the context of OpRisk modeling of extreme losses in a single risk process is that it involves compound processes. In such contexts, the extremes may be modeled at the fundamental severity process level or the annual loss compound process level. Should one wish to consider the compound process models, this involves modifications to standard EVT theory to incorporate random numbers of losses per year. We discuss this feature at the end of this chapter.

In general, the possibility of extreme event risk is present in all areas of risk management and one of the key challenges that a risk manager faces is to implement risk management models which allow for rare but damaging events and permit the measurement of their consequences. When we think of risks, we treat them as random variables, which map possible business situations or outcomes of a particular business units activity into values representing profits and

losses. In this chapter, we chose to consider the possible loss values of a risk to be suitably approximated by a statistical model which in this case will correspond to considering the losses as distributed according to a probability distribution. The particular probability distribution of the losses will never be observed exactly; however, historical losses from similar risk processes may provide some information on properties of the loss distribution.

We consider losses to be 'extreme' events when a risk process produces values from the 'tail' of its distribution. From a statistical perspective, a model for a risk process is developed by selecting a particular probability distribution. We will focus in this chapter on parametric distributional models. In general, we will aim to estimate the parameters of this distribution for the losses, which we referred to as a severity distribution in an LDA model via statistical analysis utilizing empirical data such as observed losses, expert opinion and scenario analysis.

In this section, we discuss EVT, which provides a tool which attempts to provide us with the best possible estimate of the tail area of the distribution. Importantly, for the context of OpRisk modeling of rare-high consequence loss processes, we note that even in the absence of observed loss data, EVT still provides important implications and intuition on the type of distribution one may consider to ensure that extreme risks are handled conservatively, which is an important aspect of modeling and capital estimation in Basel II and Basel III under the advanced measurement approach. We focus in these sections on the univariate EVT theory, although we note that there are also many generalizations studied in the context of multivariate extremes, see, for example, Tawn (1990), Coles & Tawn (1991), McNeil (1999) and Joe (1990).

We present fundamental results from EVT that are of direct relevance in analysis of OpRisk loss distributional models (LDA). These results will directly aid in the analysis of distributions of annual loss random variables in LDA OpRisk models especially when one considers the analysis of risk measures. The results apply directly when developing approximations of heavy-tailed LDA model quantiles, known as *single loss approximations*, which are used to approximate capital quantities such as value-at-risk (VaR) required by Basel II, Basel III and Solvency II; this will be considered in Chapters 7 and 8.

There are many excellent texts and summary papers on the background of EVT, we refer the interested reader to the texts of Beirlant *et al.* (2004), Kotz & Nadarajah (2000), Resnick (2007b), Embrechts *et al.* (1997) and Balkema *et al.* (2007) and the primer papers developed for risk modeling settings in Diebold *et al.* (2000), McNeil (1999) and Embrechts *et al.* (1999). The aim of this section is not to provide a comprehensive review of EVT, instead we provide a selected subset of results that will be of direct relevance in the development of OpRisk LDA models; these are largely based on Chapter 2 of Beirlant *et al.* (2004) and papers which present concepts from EVT in the context of risk and insurance developed by Embrechts *et al.* (1997), McNeil (1999), Reiss & Thomas (2007), Beirlant *et al.* (1996) and specifically in the context of OpRisk Chavez-Demoulin *et al.* (2006).

As discussed in McNeil (1999), in general, there are two basic kinds of model that practitioners can consider for extreme values and the motivation of their models; these are the 'Block Maxima' (BM) models and the Peaks Over Threshold (POT) models. Historically, the BM approach was first developed and involved models for the largest observed loss collected from large samples of identically distributed losses in a given period, in OpRisk a year. We begin this chapter with an introduction to the theoretical properties of this class of models, that is, a study of the limiting distribution of maximum loss which will introduce the concept of the generalized extreme value (GEV) distribution and the concept of the maximum domain of attraction. This provides a solid basis for the understanding of fundamental concepts that have been studied in the literature on EVT modeling. In addition, these results provide the theoretical underpinning for the results relating to the generalized Pareto distribution (GPD) in the

POT approach. Although the BM method is viewed as generally less practically applicable in developing EVT models, because of the manner in which the data is utilized, from a practical perspective, the BM approach to modeling extreme losses has been advocated in McNeil (1999) as a practically relevant methodology for providing estimates of stress losses.

We then proceed with a subsection discussing practical applications and estimation of EVT models, after which we switch to the more recent approaches which have come from the class of POT models which involve modeling all large observations which exceed a high threshold. This modeling approach is generally perceived to be of most practical use when developing EVT models because of the manner in which the limited loss data observations are utilized. There are two classes of model in this approach to EVT: the parametric models involving the GPD based around works of Embrechts *et al.* (1999) or the semi-parametric approaches involving Hill estimators of Hill (1975), Beirlant *et al.* (1999) and Daniélsson *et al.* (1998).

2.3 Theoretical Properties of Univariate EVT – Block Maxima and the GEV Family

In this subsection, we introduce and discuss the properties and justification of the primary theoretical results developed in EVT relating to the GEV distribution which we show arises as the natural limit distribution for normalized maximum loss in a risk process.

We first introduce some elementary definitions utilized throughout this chapter relating to distributions for the severity models considered. It will also be of importance to introduce basic convergence concepts such as uniform convergence in Definition 2.1 and convergence in distribution given in Definition 2.5.

Definition 2.1 (Uniform Convergence) *Given a sequence of functions* $\{ f_n\}_{n\geq 0}$ *which are defined on any metric space, such as real valued functions, the sequence* $\{f_n\}_{n\geq 0}$ *is said to converge on* $A \subseteq \mathbb{R}$ *to a function* f_0 *if*

$$\sup_{x\in A} | f_n(x) - f_0(x)| \to 0, \tag{2.1}$$

as $n \to \infty$.

This definition of uniform convergence for sequences of functions is particularly useful when considering convergence of density functions arising from sequences of random variables, as we see in the following discussions on convergence of distributions.

Corollary 2.1 *If the functions in the sequence* $\{ f_n\}_{n\geq 0}$ *are known to be strictly non-decreasing real-valued functions, then if* f_0 *is continuous and* $f_n(x) \to f_0(x)$ *as* $n \to \infty$ *for all* $x \in \mathbb{R}$, *then* $f_n \to f_0$ *locally uniformly such that for* $a < b$,

$$\sup_{x\in[a,b]} | f_n(x) - f_0(x)| \to 0$$

as $n \to \infty$.

For a proof of such a result see Resnick (2007, chapter 0). We make a small note regarding the notation here: *from the practical point of view, if only closed intervals are considered when looking at the function, then one can think of replacing the supremum with the max.*

Definition 2.2 (Probability Distribution Function) *A probability distribution function (the law of a random variable or cumulative distribution function), denoted by an upper script, F on $[0, \infty)$ is a non-decreasing function satisfying $F(0) = 0$, $F(\infty) = 1$, and $F(x) \leq 1$ for $x < \infty$.*

Definition 2.3 (Tail Distribution Function) *The tail of a distribution function is denoted $\overline{F}(x) = 1 - F(x)$.*

Definition 2.4 (Tail Equivalence) *Two distribution functions F and G are tail equivalent if they have the same right end point $x_* = \inf\{x : F(x) = 1\} = \inf\{x : G(x) = 1\}$ and the following limit holds*

$$\lim_{x \uparrow x_*} \frac{\overline{F}(x)}{\overline{G}(x)} = c, \tag{2.2}$$

for some constant $0 < c < \infty$.

The convergence of sequences of random variables, losses in a year for a given risk process, to some limit random variable is an important concept to understand when developing and studying properties of risk process in OpRisk modeling. These concepts formalize the idea that a sequence of unpredictable loss events can sometimes be expected to settle down into a behaviour that is essentially unchanging when a large enough number of losses are observed. For example, when we consider convergence in distribution of a risk process, what we refer to is that values in the sequence of observed losses continue to change, however, they can be described by an unchanging probability distribution.

Definition 2.5 (Weak Convergence in Distribution) *Consider a sequence of random variables $\{X_n\}$ with corresponding distribution $F_n(\cdot)$. Then X_n converges in distribution (converges weakly) to a random variable X if the distribution function of X_n converges pointwise to the distribution function of X, $F(x)$, at all points where X is continuous, that is,*

$$\lim_{n \to \infty} F_n(x) = F(x), \tag{2.3}$$

for every number $x \in \mathbb{R}$ at which F is continuous. This will be denoted by $X_n \xrightarrow{d} X$.

Remark 2.1 *The standard approach to proving the weak convergence of a sequence of random variables is to study the convergence of the corresponding characteristic functions, such as in the Lindeberg central limit theorem (CLT); see, for example, details in Jacod and Shiryaev (1987).*

Furthermore, we can also define equality in distribution according to Definition 2.6.

Definition 2.6 (Equality in Distribution) *A random variable X, with distribution and moment generating functions $F_X(x) = \Pr(X \leq x)$ and $M_X(t) = \mathbb{E}[e^{tX}]$, and a random variable Y, with distribution and moment generating function $F_Y(y) = \Pr(Y \leq y)$ and $M_Y(t) = \mathbb{E}[e^{tY}]$, are equal in distribution $(X \overset{d}{=} Y)$ if $F_X(x) = F_Y(x)$, $\forall x \in \mathbb{R}$, or equivalently when the moment generating functions exist, $M_X(t) = M_Y(t)$, $\forall t \in \mathbb{R}$.*

In certain settings, as specified by the Helly–Bray Theorem, see Billingsley (1995), one can alternatively consider studying the convergence in expectations instead of the convergence in distributions.

Theorem 2.1 (Helly–Bray) *Consider random variable $X_n \sim F_n(x)$ and random variable $X \sim F(x)$, then $X_n \xrightarrow{d} X$ if and only if for all real, bounded continuous functions $g(\cdot)$ the following holds,*

$$\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]. \tag{2.4}$$

We briefly motivate the concept of convergence of sequences of random variables according to the well-known results of the CLT and its variants. Then we discuss how the modification of the sequence of linear combinations of scaled and translated random variables to the maximum of a sequence of random variables leads us to the study of extremes and EVT.

We begin by considering a sequence of n losses from a risk process denoted by

$$\{X_1, X_2, \ldots, X_n\}.$$

The study of the distributional properties of the sum of these losses

$$Z_n = X_1 + X_2 + \cdots + X_n,$$

has been and continues to be a major area of study in statistics. In particular, one will immediately consider the well-known results related to the mean of this sequence of losses. This should bring to mind the important results related to CLT (under suitable assumptions on the properties of the severity distribution of the losses). In general, the CLT considers the sum of the losses and attempts to find constants $a_n > 0$ and b_n such that the linearly transformed random variable given by $(Z_n - b_n)/a_n$ tends in distribution to a non-degenerate distribution.

In generality, a CLT can be considered as a set of weak-convergence theories, which we present in a general manner when we introduce the family of OpRisk severity distributions from the heavy-tailed α-stable models; Chapter 4 details the properties of such models. The CLT results specify the domains of attraction of the random variables Z_n asymptotically in n, that is, they specify asymptotically the sets of attractor distribution functions of Z_n.

In the context of heavy-tailed severity models for OpRisk losses, we consider the consequences of studying the distributional properties of Z_n. In elementary statistical courses, one first discovers the well-known result that a sum of many independent and identically distributed (i.i.d.) random variables will tend to be distributed according to one of a small set of attractor distributions. This result can be extended in certain cases to also consider random variables with specific types of dependence. In particular, when the variance of the i.i.d. variables is finite, the attractor distribution is the Gaussian distribution. However, if Z_n is comprised of i.i.d. random variables with infinite variances, that is, the OpRisk losses $\{X_i\}_{i=1:n}$ are each random variables with a distribution that has power law tails decreasing according to

$$\frac{1}{|x|^{\alpha} + 1},$$

where $0 < \alpha < 2$, then the attractor distribution will be an α-stable distribution with stability parameter α. Specifically, Pareto-type distributions of the OpRisk heavy-tailed severity distributions F with infinite variance will produce non-Gaussian limits for the average loss. The reason for this is that the extreme losses that may arise from such a severity model tend to corrupt the average resulting in a different asymptotic behaviour for the attractor distributions that may arise.

In the simplest form, the CLT provides conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be

approximately Gaussian distributed. That is, one can easily show the following convergence in distribution is satisfied

$$\sqrt{n}\left(\frac{(X_1+X_2+\cdots+X_n)/n-\mathbb{E}[X]}{\sqrt{\mathbb{V}\mathrm{ar}[X]}}\right)\xrightarrow{d} Normal(0,1), \text{as } n\to\infty. \tag{2.5}$$

We note that there are several variants of the CLT which extend these results in which convergence of the mean to the Normal distribution also occurs for non-identical distributions, see more discussion on this in Chapter 4. There are also related results such as the laws of large numbers, CLT refinements such as the Berry–Essen theorem in Berry (1941), Edgeworth and saddle-point approximations and many other results that have arisen from studying properties of the average or sum of such random variables, in our context losses.

At this point, it will be useful to define the order statistics of the loss process according to Definition 2.7.

Definition 2.7 (Order Statistics) *Given losses from a risk process, which are random variables $X_1, X_2, \ldots, X_n$, we define the order statistics, denoted $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ as the random variables, obtained by sorting the values (realizations) of $X_1, X_2, \ldots, X_n$ in an increasing order.*

In this section, instead of considering the average behaviour of the sum of OpRisk losses given by Z_n we instead consider the distributional properties of the maximum loss that may arise from the OpRisk loss process in which the severity distribution is heavy tailed. We characterize the distributional properties that are known about the maximum loss. In particular, we replace the study of the random variable Z_n with consideration of the maximum, which is the n-th order statistic, defined by

$$X_{(n)}=\max\{X_1, X_2, \ldots, X_n\}. \tag{2.6}$$

Given this distribution, we note that in the case of independent observed losses, we may define the distribution of the maximal loss according to the following,

$$\mathbb{P}\mathrm{r}\left[X_{(n)}\leq x\right]=\mathbb{P}\mathrm{r}\left[\cap_{i=1}^{n}(X_i\leq x)\right]=\prod_{i=1}^{n}\mathbb{P}\mathrm{r}\left[X_i\leq x\right]=(F(x))^n. \tag{2.7}$$

We address the probabilistic question of what is the limiting distribution of the maximum loss $X_{(n)}$, which we denote by $M(x)$. That is, suppose that we have observed losses $\{X_1(t), X_2(t), ...\}$ in some blocks (e.g. annual years). Furthermore, assume that we have considered these losses as independent identically distributed losses with severity distribution function F, then if we study the behaviour of the standardized maximal loss, we are interested to know under what circumstances the distribution of this maximal loss converges to some limiting distribution M as the block size increases. We will demonstrate that such a convergence will occur when the severity distribution from which the losses were realizations is in the maximum domain of attraction of M. We will show that when M is non-degenerate, then it will be given by the GEV distribution. This is a profound result in the study of extremes and can be considered as fundamental to the study of maxima as the CLT is for the study of sums or averages. This result is known as the *Fisher–Tippett–Gnedenko theorem*.

In addition, we note that many of the results we demonstrate can be extended to also addressing related questions of extremes such as the average of the k largest losses or the sum of excess losses above a specified threshold, that is, the loss amount u (when normalized by

the number of losses, this is the empirical estimator of the mean excess). However, we focus primarily on the basic question related to the maximal loss in this section.

As an introduction that only requires basic probability, we consider what happens to the distribution of the random variable $X_{(n)}$ formed from the maximal loss as the number of observed losses $n \to \infty$ when we do not apply a standardization. It will be shown in Example 2.1 that in this case, the limiting distribution of the maximal loss is always degenerate.

EXAMPLE 2.1

Consider a loss process with loss random variables $\{X_1, X_2, \ldots, X_n\}$ which are i.i.d. according to a severity distribution F. Furthermore, define the right end point of the loss distribution to be $x_* = \inf\{x : F(x) = 1\}$. Now consider what happens to the maximum loss (n-th order statistic in n samples) as the number of samples $n \to \infty$,

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n) \to x_*. \tag{2.8}$$

In addition, we have (for this i.i.d. loss process setting) the result in Equation 2.7, which provides the finite sample distribution of the maximal loss $X_{(n)}$ according to

$$\mathbb{P}\mathrm{r}[X_{(n)} < x] = \prod_{i=1}^{n} \mathbb{P}\mathrm{r}[X_i \leq x] = (F(x))^n. \tag{2.9}$$

It is now clear that we have the limiting degenerate result for the limiting distribution of the maximal loss when standardization is not applied, which is given by

$$F^n(x) \to \begin{cases} 0, & \text{for } x < x_*, \\ 1, & \text{for } x \geq x_*, \end{cases} \tag{2.10}$$

as $n \to \infty$. Therefore, the limiting sequence of $\lim_{n\to\infty} F^n(x)$ is degenerate. ■

Formalizing this, at a fundamental level, the main concern of EVT when applied to OpRisk is to search for distributions of the loss X for which one may find a sequence of constants $\{a_n; n > 1, a_n > 0\}$ and $\{b_n; n > 1\}$ such that for all values of $x \in \mathbb{R}$ for which the limit exists, the following holds asymptotically in $n \to \infty$

$$\mathbb{P}\mathrm{r}\left[\frac{X_{(n)} - b_n}{a_n} \leq x\right] \to M(x). \tag{2.11}$$

In particular, one considers scenarios in which the limiting distribution $M(x)$ of the linearly transformed maximum loss is non-degenerate. In this setting, analogous to the questions posed when studying the CLT results in elementary statistics courses, we can pose similar questions relating to the properties of the distribution of $X_{(n)}$. As stated in Beirlant *et al.* (2004), the basic goals of EVT are summarized according to the following.

1. *Extremal Limit Problem*. Determine all (non-degenerate) distributions $M(x)$ that can be obtained as the limit of Equation 2.11, see Fisher & Tippett (1928), Gnedenko (1943) and Balkema & De Haan (1972).

2. *Domain of Attraction Problem*. Characterize the distributions of the OpRisk severity models F for which there exist sequences $\{a_n; n \geq 1\}$ and $\{b_n; n \geq 1\}$ such that the limit in Equation 2.11 is satisfied for any such specific limit distributions $M(x)$.
3. *Second-Order Problem*. Given a loss process with distribution F for which there exists a sequence of a_n and b_n standardizing constants such that F is in the domain of attraction of M and we know explicitly the distribution M, at what rate will this convergence occur?

In considering the 'domain of attraction problem', one is considering a limiting distribution $M(x)$ for the sequence-transformed maximum losses, $\{a_n^{-1}(X_{(n)} - b_n)\}_{n\geq 1}$, and inquiring about the necessary and sufficient conditions on the possible heavy-tailed severity distributions of the losses, $F \in \mathscr{D}(M)$, such that the limiting distribution $M(x)$ is obtained. That is, one studies for the limiting distribution $M(x)$ the domain of attraction, denoted by $\mathscr{D}(M)$, that is the subset of possible heavy-tailed severity distribution models that will result in loss processes for which the maximum loss achieved in a year will have the limiting distribution $M(x)$ in Equation 2.11.

In addition, it can be shown that all risk processes with losses $X_i \sim F$, where we assume that the severity distribution F is both heavy tailed and continuous and strictly increasing, then the solution to the 'extremal limit problem' satisfies the results in Definition 2.8.

Definition 2.8 (Extremal Limit Problem Solution) *For heavy-tailed severity distribution functions $F(x)$ which are continuous and strictly increasing, the limiting distribution of the sequence of linearly transformed maximal losses in a year that satisfies the Helly–Bray theorem,*

$$\mathbb{E}\left[g\left(a_n^{-1}(X_{(n)} - b_n)\right)\right] = n \int_{-\infty}^{\infty} g\left(\frac{x - b_n}{a_n}\right) F^{n-1}(x)\, dF(x), \tag{2.12}$$

for real, bounded and continuous functions $g(\cdot)$, is of the form given by

$$M(x;\gamma) = \begin{cases} \exp\left(-(1+\gamma x)^{-1/\gamma}\right), & \text{for } 1+\gamma x > 0, \\ \exp(-\exp(-x)), & \text{for } x > 0, \gamma = 0. \end{cases} \tag{2.13}$$

It is parameterized by the extreme value index (EVI) parameter $\gamma \in \mathbb{R}$, which plays the role of distinguishing three important sub-classes of family of limiting distributions from the Frechet–Pareto type ($\gamma > 0$), the Gumbel type ($\gamma = 0$) or the Weibull type ($\gamma < 0$).

It is also useful to note the following definition for distributions of the same type, or within a family, given in Definition 2.9. In addition, notation for the inverse distribution function or quantile function are provided along with the specification of the tail quantile function, see Definitions 2.12–2.14.

Definition 2.9 (Distributions of Common Type) *Distributions of the same type will be defined as those that differ only in location and scale.*

Using this definition of distributions of the same type, one can show that any EVT distribution can be represented according to the following distribution known generically as the GEV distribution that we denote by H and specified in Definition 2.10.

Definition 2.10 (Generalized Extreme Value (GEV) Distribution) *We consider distribution M of type H to be defined such that for some $a > 0$ and $b \in \mathbb{R}$, we can define*

$$M(x;\gamma) = H_\gamma\left(\frac{x-b}{a}\right), x \in \mathbb{R}, \tag{2.14}$$

with M also given by Equation 2.13. More broadly, we can summarize all the possible limiting distributions according to the GEV distribution, which is defined by setting $a = \mu$ and $b = \sigma$ to obtain

$$\mathbb{P}\mathrm{r}[X < x;\mu,\sigma,\gamma] = \exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\}, \tag{2.15}$$

for $1+\gamma(x-\mu)/\sigma > 0$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the shape parameter. Furthermore, the density function is given by

$$f(x;\mu,\sigma,\gamma) = \frac{1}{\sigma}\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\gamma)-1}\exp\left\{-\left[1+\gamma\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\gamma}\right\} \tag{2.16}$$

again for $1+\gamma(x-\mu)/\sigma > 0$. In addition, the support of a random variable $X \sim H_\gamma\left(\frac{x-\mu}{\sigma}\right)$ is given by

$$S_X = Supp\left(H_\gamma\left(\frac{x-\mu}{\sigma}\right)\right) = \begin{cases}\left[\mu-\frac{\sigma}{\gamma},\infty\right], & \gamma > 0,\\ [-\infty,\infty], & \gamma = 0,\\ \left[-\infty,\mu-\frac{\sigma}{\gamma}\right], & \gamma < 0.\end{cases} \tag{2.17}$$

In Figure 2.1, we provide some example plots of the density functions for different parameter ranges from this family of GEV distributions. We begin with the EVT distribution in which $\gamma = 1$, and we consider a range of values $H_{\gamma;\mu,\sigma}$. Then we fix $\mu = 0$ and $\sigma = 1$ and study the influence of the tail index in the GEV family in Figure 2.2.

In addition, it is often useful for estimation of the GEV model parameters to have expressions for the moments and summary statistics of this distribution with respect to the model parameters, these are provided in Definition 2.11.

Definition 2.11 *The moments and summary statistics of the GEV distribution are known in closed form with respect to the model parameters and given according to*

$$\mathbb{E}[X] = \begin{cases}\mu+\sigma\dfrac{\Gamma(1-\gamma)-1}{\gamma}, & \gamma < 1, \gamma \neq 0,\\ \mu+\sigma e_C, & \gamma = 0,\\ \infty, & \gamma \geq 1,\end{cases}$$

$$\mathbb{V}\mathrm{ar}[X] = \begin{cases}\dfrac{\sigma^2}{\gamma^2}\left(\Gamma(1-2\gamma)-\Gamma(1-\gamma)^2\right), & \gamma < \dfrac{1}{2}, \gamma \neq 0,\\ \dfrac{\sigma^2\pi^2}{6}, & \gamma = 0,\\ \infty, & \gamma \geq \frac{1}{2},\end{cases}$$

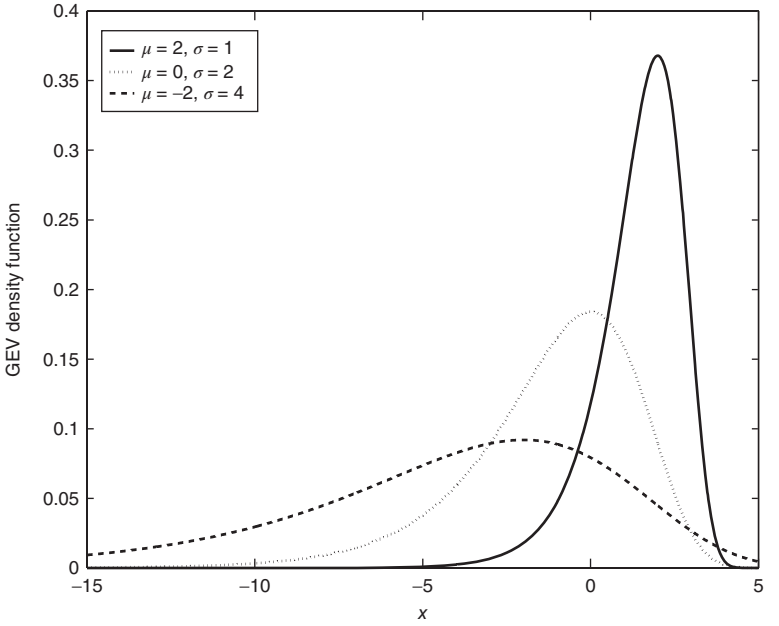

FIGURE 2.1 EVT density plots of $H_{1;\mu,\sigma}$ for different location and scale settings.

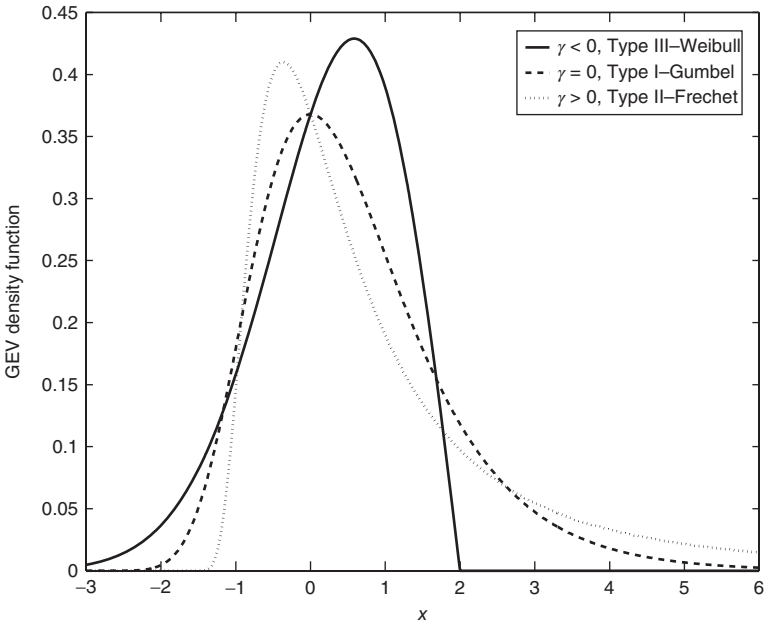

FIGURE 2.2 GEV density plots of $H_{\gamma;0,1}$ for the settings: Frechet $\gamma = 1/2$; Gumbel $\gamma = 0$ and Weibull $\gamma = -1/2$.

$$\text{mode}[X] = \begin{cases} \mu + \dfrac{\sigma}{\xi}[(1+\gamma)^{-\gamma} - 1], & \gamma \neq 0, \\ \mu, & \gamma = 0, \end{cases}$$

$$\mathbb{S}\text{kewness}[X] = \begin{cases} \dfrac{\Gamma(1-3\gamma) - 3\Gamma(1-\gamma)\Gamma(1-2\gamma) + 2\Gamma(1-\gamma)^3}{(\Gamma(1-2\gamma) - \Gamma(1-\gamma)^2)^{3/2}}, & \gamma \neq 0, \\ \dfrac{12\sqrt{6}\zeta(3)}{\pi^3}, & \gamma = 0, \end{cases}$$

$$\mathbb{K}\text{urtosis}[X] = \begin{cases} \frac{\Gamma(1-4\gamma) - 4\Gamma(1-\gamma)\Gamma(1-3\gamma) + 6\Gamma(1-2\gamma)\Gamma(1-\gamma)^2 - 3\Gamma(1-\gamma)^4}{(\Gamma(1-2\gamma) - \Gamma(1-\gamma)^2)^2} - 3, & \gamma \neq 0, \\ \dfrac{12}{5}, & \gamma = 0. \end{cases}$$

where e_C denotes Euler's constant given by

$$e_C = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln\, n \right) \tag{2.18}$$

and $\zeta(x)$ is the Riemann zeta function.

To proceed with analysis of results for EVT, it will be useful to define a few properties of distribution functions that will be utilized. This will involve the definition of a pseudo-inverse of a function, a tail quantile function and associated properties of these functions of relevance to OpRisk practitioners seeking to gain insight into the basic results of EVT and their applicability to OpRisk modeling.

A real-valued, continuous and strictly monotone univariate function will possess an inverse on its range. This is also generalized if one considers relaxing the conditions of continuity and strict monotonicity by working with the generalized inverse in Definition 2.12. This concept will be of direct importance to OpRisk when one considers quantile functions which are often used in evaluation of risk measures for regulatory reporting. A review of the basic properties of generalized inverses can be found in Embrechts and Hofert (2010) and Resnick (2007).

Definition 2.12 (Generalized Inverse) *Given an increasing function $F : \mathbb{R} \to \mathbb{R}$ with $F(-\infty) = \lim_{x\downarrow -\infty} F(x)$ and $F(\infty) = \lim_{x\uparrow\infty} F(x)$, the generalized inverse is given by $F^{\leftarrow} : \mathbb{R} \to \bar{\mathbb{R}} = [-\infty, \infty]$ according to*

$$F^{\leftarrow}(y) = \inf\{x \in \mathbb{R} : F(x) \geq y\},\ y \in \mathbb{R}. \tag{2.19}$$

In addition, the following properties of generalized inverses will be of relevance to considering the relationship between a quantile function and a distribution function; their proofs are given in Embrechts & Hofert (2010, proposition 2.1).

Proposition 2.1 *The following basic properties of generalized inverse functions apply for $x, y \in \mathbb{R}$.*

1. *$F^{\leftarrow}(y) = -\infty$, if and only if $F(x) \geq y$ for all $x \in \mathbb{R}$, and $F^{\leftarrow}(y) = \infty$, if and only if $F(x) < y$, for all $x \in \mathbb{R}$.*
2. *$F^{\leftarrow}$ is increasing and if $F^{\leftarrow}(y) \in (-\infty, \infty)$, $F^{\leftarrow}$ is left continuous at y and admits a limit from the right of y.*

3. $F^{\leftarrow}(F(x)) \leq x$. *If* F *is strictly increasing,* $F^{\leftarrow}(F(x)) = x$.
4. *If* F *is right continuous, then* $F^{\leftarrow}(y) < \infty$ *implies that* $F\left(F^{\leftarrow}(y)\right) \geq y$. *In addition, if* $y \in ranF \cup \{\inf\ ranF, \sup\ ranF\}$, *then* $F\left(F^{\leftarrow}(y)\right) = y$. *Also, if* $y < \inf\ ranF$, *then* $F\left(F^{\leftarrow}(y)\right) > y$, *and if* $y > \sup\ ranF$, *then* $F\left(F^{\leftarrow}(y)\right) < y$.
5. $F(y) \geq y$ *implies that* $x \geq F^{\leftarrow}(y)$.
6. F *is continuous if and only if* $F^{\leftarrow}$ *is strictly increasing on* $\mathbb{R}$. F *is strictly increasing if and only if* $F^{\leftarrow}$ *is continuous on* $ranF$.
7. *Finally, if* F_1 *and* F_2 *are right continuous transformations with the same properties as* F, *then the composite function* $(F_1 \circ F_2)^{\leftarrow} = F_1^{\leftarrow} \circ F_2^{\leftarrow}$.

This general Definition 2.12 of an inverse can now be considered in the case in which we have a loss process with a severity distribution function F, to give the definition of the quantile function.

Definition 2.13 (Quantile Function) *Given an OpRisk severity model in which the distribution* F *is considered continuous, we define the quantile function, according to the inverse function given by* $F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$, *which we will also refer to as the quantile function* $Q(y)$, *with* $y \in [0, 1]$.

Definition 2.14 (Tail Quantile Function) *Given an OpRisk severity model in which the distribution* F *is considered continuous, we define the quantile function, according to the inverse function given by* $F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$ *we will refer to the tail quantile function as* $U(y) = F^{\leftarrow}\big(1 - 1/y\big)$, *with* $y \in [0, 1]$.

Remark 2.2 *Note, we will also use* $Q_X(y)$ *and* $U_X(y)$ *as notation for quantile function and tail quantile when we need to make explicit the random variable* X *for which the quantile is defined.*

In Figure 2.3, we provide an example that illustrates the need to consider generalized inverses when working with distribution functions. We can see that the standard definition of an inverse would not allow us to consider definitions for quantile functions and tail quantile functions if the cumulative distribution functions (cdfs) we consider have atoms, i.e. they are of mixed type.

In addition, the following relationships given in Proposition 2.2 between a distribution function and a tail quantile function will be of relevance throughout this chapter and are based on properties defined in Proposition 2.1 of generalized inverses when applied to distribution functions and quantile functions.

Proposition 2.2 *The following relationships hold between a distribution function and its tail quantile function, for* $y \in [0, 1]$.

1. *The tail quantile relates to the quantile function of distribution* F *via* $U(y) = Q(1 - 1/y)$.
2. *If* $z < U(y)$, *then* $1 - F(z) > 1/y$.
3. *For all* $y > 0$, *then* $1 - F(U(y)) \leq 1/y$.
4. *For all* $x < \inf\{x : F(x) = 1\}$, *then* $U\big((1 - F(x))^{-1}\big) < x$.
5. *If the distribution function* F *is strictly continuous, then* $1 - F(U(y)) = 1/y$.
6. *If the tail quantiles function* U *is strictly continuous, then* $U\left((1 - F(x))^{-1}\right) = x$.

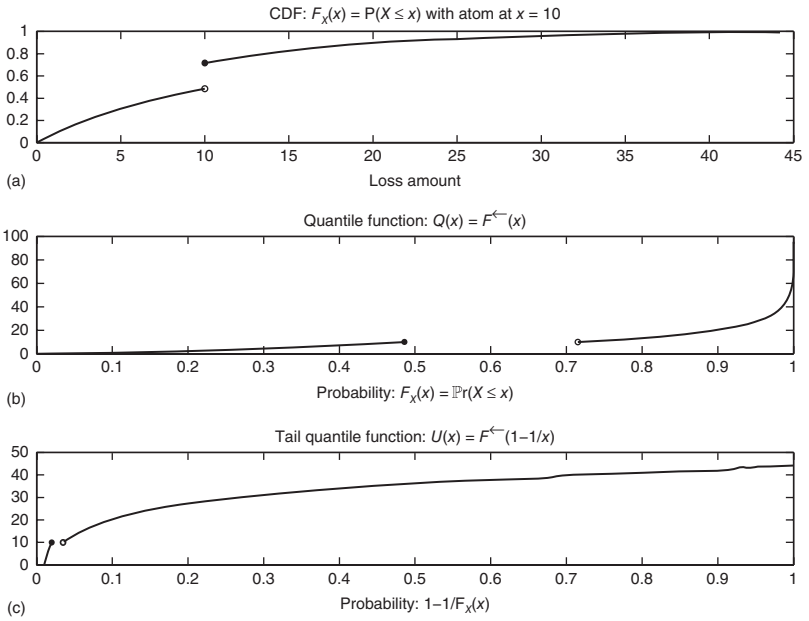

FIGURE 2.3 **In this example, we consider an arbitrary cdf function for a severity distribution with an atom at $x = 10$ corresponding to a discontinuity in a mixed-type cdf $F_X(x)$. (a) Plot of cdf; (b) plot of quantile function; (c) plot of tail quantile function.**

7. *If the distribution function F admits a density function f on its domain of definition, then one has the relationship*

$$1 - F(U(y)) = \int_{U(y)}^{inf\{x:F(x)=1\}} f(s)\,ds = \frac{1}{y} \tag{2.20}$$

and, in addition, the quantile function U will also admit a derivative u and the following relationship holds

$$f(U(y))\,u(y) = y^{-2}. \tag{2.21}$$

As a consequence, we note the following relationships between Definitions 2.13 and 2.14, which is utilized throughout this chapter,

$$U(y) = Q\left(1 - \frac{1}{y}\right) = x, \tag{2.22}$$

and we also have the relationship for the distribution function at the y-th quantile, for $y \in [0, 1]$,

$$F(x) = 1 - \frac{1}{y}. \tag{2.23}$$

We now return to discussion on the GEV model. Completing the specifications of the quantile function and its properties, one can show that the quantile function of the GEV distribution is given by the expression in Definition 2.15.

Definition 2.15 (GEV Quantile Function) *Given an OpRisk severity model in which the distribution $F \in \mathscr{D}(M)$ is in the domain of attraction of the GEV distribution, then the quantile function can be obtained as a function of the EVI according to*

$$Q(p;\gamma) = \frac{\left(\frac{1}{\ln(1/p)}\right)^{\gamma} - 1}{\gamma}, \; p \in (0,1), \gamma \neq 0, \tag{2.24}$$

and in the Gumbel case

$$Q(p;0) = \ln\left(\frac{1}{\ln(1/p)}\right), \; p \in (0,1), \gamma = 0. \tag{2.25}$$

In Figure 2.4, we plot the quantile function for the GEV from each of the three types for a range of EVI values.

Now, considering the transformed maximum loss given by the random variable

$$a_n^{-1}(X_{(n)} - b_n)$$

and applying the Helly–Bray Theorem 2.1, one can show that this sequence of random variables representing the standardized maximum loss will converge weakly to X as $n \to \infty$ when the following condition for all real, bounded and continuous functions $g(\cdot)$ is satisfied,

$$\mathbb{E}\left[g\left(a_n^{-1}\left(X_{(n)} - b_n\right)\right)\right] \to \int_{-\infty}^{\infty} g(\nu)\, dM(\nu), \tag{2.26}$$

as $n \to \infty$ and $M(\nu) = \mathbb{P}\mathrm{r}\left[a_n^{-1}\left(X_{(n)} - b_n\right) \leq \nu\right]$.

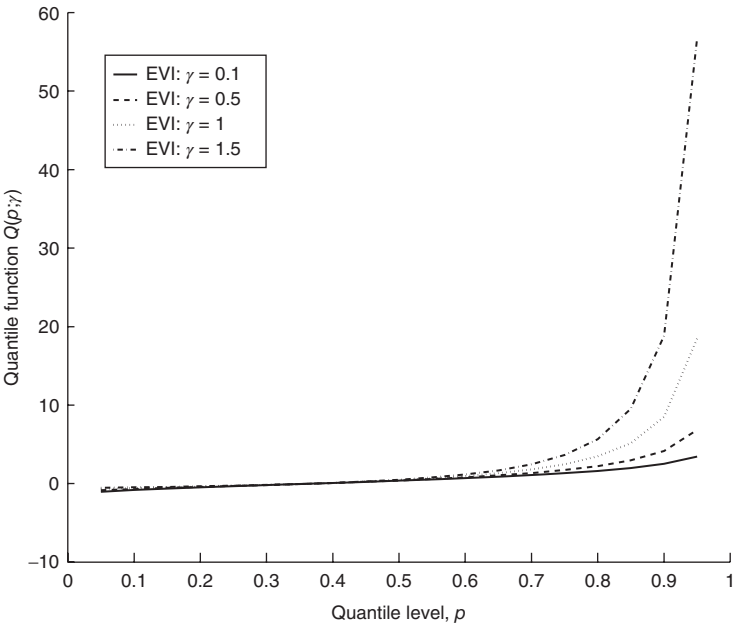

FIGURE 2.4 **Quantile function plotted for a range of EVI values for the Frechet–Pareto family of the GEV distribution.**

Following the approach of Beirlant *et al.* (2004), one may proceed with an understanding of how this result is obtained. First rewrite the left-hand side expectation for the scaled and translated maxima in terms of the equivalent expression defined in Equation 2.7. One obtains the equivalent condition on the expected value of function $g(\cdot)$ of the standardized maximal loss given by

$$\mathbb{E}\left[g\left(a_n^{-1}\left(X_{(n)}-b_n\right)\right)\right]=n\int_{-\infty}^{\infty}g\left(\frac{x-b_n}{a_n}\right)F^{n-1}(x)dF(x). \tag{2.27}$$

Next, one may substitute the definition of the tail quantile function, discussed in Definition 2.14, into Equation 2.27 to obtain

$$\mathbb{E}[g(a_n^{-1}(X_{(n)}-b_n))]=\int_0^n g\left(\frac{U\left(\frac{n}{\nu}\right)-b_n}{a_n}\right)\left(1-\frac{\nu}{n}\right)^{n-1}d\nu. \tag{2.28}$$

As pointed out in Beirlant *et al.* (2004), one can clearly see that in the limit as $n\to\infty$, the integral extends to cover a domain of $\mathbb{R}^+$ and the second term in the integrand $(1-\nu/n)^{n-1}\to\exp(-\nu)$. Hence, one concludes that the limit of $\mathbb{E}[g(a_n^{-1}(X_{(n)}-b_n))]$ exists when a sequence a_n can be obtained, which will make $a_n^{-1}(X_{(n)}-b_n)$ convergent for all positive ν, see Figure 2.5.

Next we wish to consider what happens to our representation of $\mathbb{E}[g(a_n^{-1}(X_{(n)}-b_n))]$ in Equation 2.27 as we observe an increasingly larger number of losses from the risk process, $n\to\infty$. In other words, we will now consider the behaviour of the argument of the function $g(\cdot)$ as we observe an increasing number of losses. If we can show that this argument, comprised of $a_n^{-1}\big(U(n/\nu)-b_n\big)$, converges for all $\nu>0$, then we can conclude that the resulting

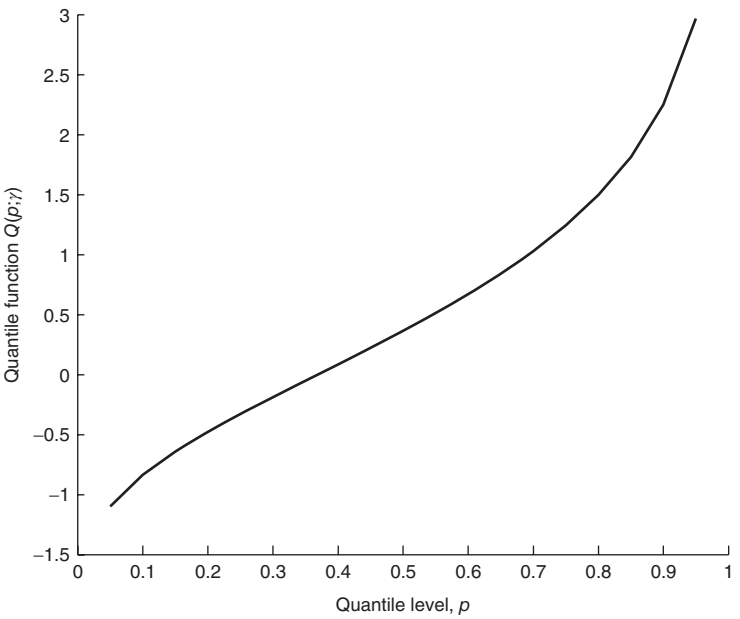

FIGURE 2.5 Quantile function plotted for the Gumbel family of the GEV distribution.

expectation $\mathbb{E}[g(a_n^{-1}(X_{(n)} - b_n))]$ will satisfy Equation 2.26 and, therefore, that the convergence originally discussed relating to the distribution of the standardized maximal loss in Equation 2.11 will have a limiting distribution $M(x)$.

Without the loss of generality, we set $b_n = U(n)$, as suggested in Beirlant *et al.* (2004), and, therefore, it is apparent that to ensure the convergence, it suffices to consider under which circumstances the sequence a_n will result in existence of a limit for $a_n^{-1}(U(n/\nu) - U(n))$. It is now apparent that this convergence will be obtained, should we impose the following condition in Definition 2.16. This definition is known as the *extremal domain of attraction condition* and in presenting it we adopt the notation of Beirlant *et al.* (2004) in the specification of this condition, as it serves to highlight the fact that all possible limits are described by a one-parameter family of functions $h_\gamma(\cdot)$ defined below.

Definition 2.16 (Extremal Domain of Attraction (Condition D_I)) *The extremal domain of attraction for severity distributions $F \in \mathscr{D}(M)$ is specified by a one-parameter family of functions $h_\gamma(\cdot)$, characterizing the class of distributions that have tail quantile functions satisfying*

$$h_\gamma(u) := \lim_{x \to \infty} \frac{U(xu) - U(x)}{a(x)}, \tag{2.29}$$

for a positive function $a(x)$ and $u > 0$.

For the derivation of the result which shows that all possible limits will be described by a single-parameter function $h_\gamma(u)$, the interested reader is referred to (Beirlant *et al.* 2004, pp. 48–51).

We illustrate this extremal domain of attraction condition for the GEV distribution M with EVIγ by considering Condition D_I, which specifies conditions based on the tail distribution function $U(x)$ and the function $a(x)$. In Example 2.2, we illustrate examples of convergence under conditions on the sequence of centering and scaling functions a_n and $b_n = U(n)$ to obtain each of the GEV sub-families. Example 2.2 below we show the examples of convergence to the Frechet–Pareto type ($\gamma > 0$), and the Gumbel type ($\gamma = 0$).

EXAMPLE 2.2 Effect of Sample Size on Convergence Rate.

This example illustrates, for simple key members of the three GEV sub-families, empirical rates of convergence of the distribution of the extreme as a function of the sample size for a range of quantile values. We consider the general starting point in which we defined the sequence of random variables and their resulting weak convergence in distribution to the GEV. In each case, we set $b_n = U(n)$ and then consider particular choices for the function a_n:

1. In the Frechet–Pareto case, we set $a_n = a(n) = \gamma x^\gamma$ because we will consider the Pareto distribution example in which the tail quantile function is given by $U(x) = x^\gamma$ giving

$$F_{X_{(n)}}\left(\frac{x_{(n)} - b_n}{a_n}\right) = \Pr\left[\frac{X_{(n)} - U(n)}{\gamma U(n)} \leq x\right] = \left(1 - \frac{(1+x)^{-1}}{n}\right)^n \to \exp(-(1+x)^{-1}), \text{as } n \to \infty. \tag{2.30}$$

In Figure 2.6, we present the empirical analysis of the rate of convergence of the Pareto distribution example for $n = 2, 5, 10, 15$ versus the limiting case.

2. In the Gumbel case, we set $a_n = a(n) = 1/\lambda$ because we will consider the exponential distribution in which the tail quantile function is given by

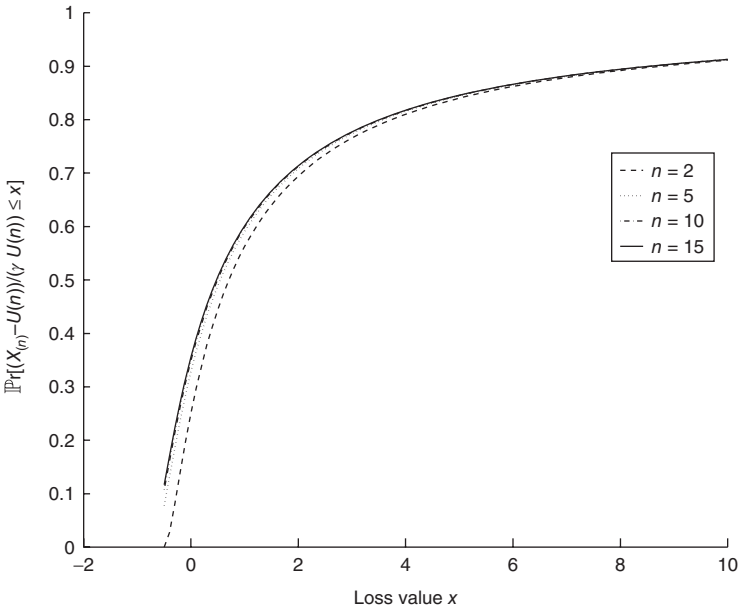

FIGURE 2.6 **Ideal scenario: convergence rate of strict Pareto distribution to extremal limit as a function of sample size n.**

$U(x) = \lambda^{-1} \ln\, x$ giving

$$F_{X_{(n)}}\left(\frac{x_{(n)} - b_n}{a_n}\right) = \mathbb{P}\mathrm{r}\left[\frac{X_{(n)} - U(n)}{1/\lambda} \leq x\right] = \left(1 - \frac{\exp(-x)}{n}\right)^n \rightarrow \exp(-\exp(-x)), \text{as } n \rightarrow \infty. \tag{2.31}$$

In Figure 2.7, we present the empirical analysis of the rate of convergence of the exponential distribution example for $n = 2, 5, 10, 15$ versus the limiting case.

3. In the Weibull case, one may set $a_n = a(x) = -\gamma(x_* - U(x))$ so that one may then consider the example generalizing the uniform distribution in which the tail quantile function is given by $U(x) = x_*(1 - x^{1/\beta})$ on domain $[1, \infty)$ giving

$$F_{X_{(n)}}\left(\frac{x_{(n)} - b_n}{a_n}\right) = \mathbb{P}\mathrm{r}\left[-\frac{X_{(n)} - U(n)}{\gamma(x_* - U(n))} \leq x\right] = \left(1 + \frac{x-1}{n}\right)^n \rightarrow \exp(-(1-x)), \text{as } n \rightarrow \infty. \tag{2.32}$$

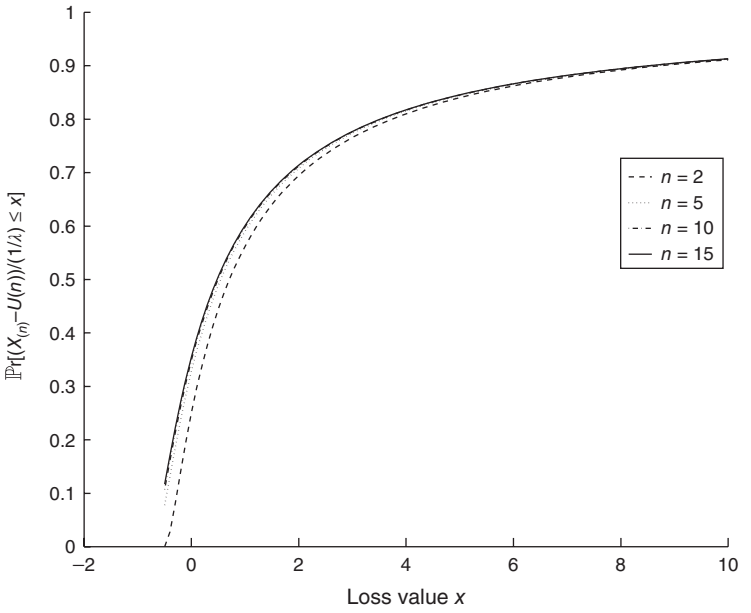

FIGURE 2.7 **Ideal scenario: convergence rate of exponential distribution to extremal limit as a function of sample size n.**

In addition, it is worth mentioning at this point the following related conditions, presented in Proposition 2.3, to that which was described by the extremal domain of attraction (Condition D_I). These conditions are often relevant from a practical perspective when considering continuous severity distribution models, see detailed discussion in (Beirlant *et al.* 2004, propositions 2.1 and 2.2).

Proposition 2.3 *A severity distribution F is a member of the extremal domain of attraction $F \in \mathscr{D}(M)$ if and only if one of the following conditions is satisfied:*

1. *The heavy-tailed severity distribution F for the OpRisk losses has a tail quantile function $U(y) = F^{\leftarrow}(1 - 1/y)$ satisfying the extremal domain of attraction condition, see Definition 2.16 (Condition D_I);*
2. *Or, (Condition D_{II}) there exists an auxiliary function $b(\cdot)$, the EVI satisfies $1 + \gamma\nu > 0$ and the severity distribution of the risk process F, satisfies the following limit*

$$\lim_{y \to \inf\{x:F(x)=1\}} \frac{1 - F(y + b(y)\nu)}{1 - F(y)} \to (1 + \gamma\nu)^{-1/\gamma} =: \eta_{\gamma}(\nu); \tag{2.33}$$

3. *Or, (Condition D_{III}) there exists a positive sequence b_n and a positive, non-constant function H, such that the heavy-tailed severity distribution F and its associated tail quantile function U for the OpRisk losses satisfy the following limit*

$$\lim_{n \to \infty} n\{1 - F(U(n) + b_n\nu)\} \to H(\nu); \tag{2.34}$$

4. *Or, Condition D_{IV}, which is a necessary condition for a distribution F to satisfy in order for Condition D_I to hold, which is satisfied when*

$$\lim_{x\to\infty} \frac{U(x)}{a(x)}\{\ln U(xu) - \ln U(x)\} \to \begin{cases} \ln(u), & \text{if } \gamma \geq 0, \\ \dfrac{u^\gamma - 1}{\gamma}, & \text{if } \gamma < 0 \text{ and } \inf\{x : F(x) = 1\} > 0. \end{cases} \tag{2.35}$$

The proof of these results are detailed in Beirlant *et al.* (2004, section 2.9.4). We complete the specification of these technical conditions for the extremal domain of attraction, with a brief discussion on the basic relationships between them, given in Corollaries 2.2 and 2.3.

Corollary 2.2 *If the severity distribution F satisfies Condition D_{II}, then one can show that the auxiliary function $b(\cdot)$ will satisfy the limit,*

$$\lim_{y\to\inf\{x:F(x)=1\}} \frac{b(y + \nu b(y))}{b(y)} \to u^\gamma = 1 + \gamma\nu. \tag{2.36}$$

Furthermore, the auxiliary function b can be represented by

$$b(y) = a(U^{\leftarrow}(y)). \tag{2.37}$$

Corollary 2.3 *If the severity distribution F satisfies Condition D_{IV}, then one can show that the following limit of the tail quantile function holds,*

$$\lim_{x\to\infty} \frac{U(x)\mathbb{E}\left[\ln\frac{X}{U(x)}\middle| X > U(x)\right]}{a(x)} \to \begin{cases} 1, & \text{if } \gamma \geq 0, \\ \int_1^\infty \dfrac{u^\gamma - 1}{\gamma}\dfrac{du}{u^2} = \dfrac{1}{1-\gamma}, & \text{if } \gamma < 0 \text{ and } \inf\{x : F(x) = 1\} > 0. \end{cases} \tag{2.38}$$

Furthermore, one can also make an explicit representation of the auxiliary function $b(\cdot)$ in Condition D_{II} can be represented by the function

$$b(t) = \begin{cases} t\mathbb{E}\left[\ln(X/t)|X > t\right], & \text{if } \gamma \geq 0, \\ t(1-\gamma)\mathbb{E}\left[\ln(X/t)|X > t\right], & \text{if } \gamma < 0, \end{cases} \tag{2.39}$$

and standardizing function $a(x)$ can be represented by

$$a(x) = \begin{cases} U(x)\mathbb{E}\left[\ln(X/U(x))|X > U(x)\right], & \text{if } \gamma \geq 0, \\ (1-\gamma)U(x)\mathbb{E}\left[\ln(X/U(x))|X > U(x)\right], & \text{if } \gamma < 0. \end{cases} \tag{2.40}$$

Returning to the derivation of the result in Definition 2.8, we utilize the result derived in Equation 2.27, and assuming the severity distribution considered satisfies the imposed condition on the extremal domain of attraction in Definition 2.16 (Condition D_I), then one may obtain a limiting expression for the right-hand side of Equation 2.26 given by

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left[g\left(a_n^{-1}(X_{(n)}-b_n)\right)\right] &= \lim_{n\to\infty}\int_0^n g\left(\frac{U(n/\nu)-b_n}{a_n}\right)\left(1-\frac{\nu}{n}\right)^{n-1}d\nu \\ &= \int_0^\infty g\left(h_\gamma(1/\nu)\right)\exp(-\nu)d\nu =: \int_{-\infty}^\infty g(u)\,dM(u;\gamma).\end{aligned} \tag{2.41}$$

We may summarize the technical domain of convergence conditions specified in this section for either the tail quantile function or the distribution function of the severity model according to Table 2.1.

TABLE 2.1 Summary of conditions on the severity distribution function F and the severity tail quantile function U to ensure $F \in \mathscr{D}(M)$.

EVI Range	Conditions on tail quantile U	Conditions on distribution F
General case	Condition D_I	Condition D_{II}
$\gamma \in \mathbb{R}$	$\lim_{x\to\infty}\frac{U(ux)-U(x)}{a(x)} = h_\gamma(u)$	$\lim_{y\to\inf\{x:F(x)=1\}}\frac{1-F(y+b(y)\nu)}{1-F(y)} = \eta_\gamma(\nu)$ with $b(y)$ given in Equation 2.39
	Condition D_{IV} $\lim_{x\to\infty}\frac{U(x)}{a(x)}\ln\left(\frac{U(ux)}{U(x)}\right) = \begin{cases}\ln(u), & \gamma \geq 0,\\ \frac{u^\gamma-1}{\gamma}, & \gamma<0.\end{cases}$ with $a(x)$ given in Equation 2.40	
$\gamma > 0$	$\lim_{x\to\infty}\frac{U(ux)}{U(x)} = u^\gamma,\ u>0$	$\lim_{y\to\infty}\frac{1-F(y\nu)}{1-F(y)} = \nu^{-1/\gamma},\ \nu>0$

We may now summarize these results pertaining to the GEV approach to EVT, in the OpRisk setting that we consider, according to the result in Theorem 2.2. Additional results proving the uniqueness of these limiting distributional results are provided in Gnedenko (1943), De Haan & Resnick (1996) and Beirlant *et al.* (2004).

Theorem 2.2 *Consider a severity distribution $F \in \mathscr{D}(M)$ that satisfies the extremal domain of attraction condition and a sequence of loss random variables in a given calendar year t, denoted by $\{X_1(t), \ldots, X_n(t)\}$, then for $\gamma \in \mathbb{R}$ the following resulting extremal distribution types are attainable, as a function of the sign of γ:*

1. *if $\gamma > 0$, one has the Frechet–Pareto case characterized by*

$$\mathbb{E}\left[g\left(a_n^{-1}\left(X_{(n)}-b_n\right)\right)\right] \to \int_{-\gamma^{-1}}^\infty g(u)d(\exp(-(1+\gamma u)^{-1/\gamma})). \tag{2.42}$$

In addition, one can show that the ratio of the standardizing function to the tail quantile function converges to the EVI,

$$\lim_{x\to\infty}\frac{a(x)}{U(x)}\to\gamma, \tag{2.43}$$

and, furthermore, that $U(x)$ is of the same regular variation as standardizing function a.

2. *if $\gamma=0$, one has the Gumbel case characterized by*

$$\mathbb{E}\left[g\left(a_n^{-1}(X_{(n)}-b_n)\right)\right]\to\int_{-\infty}^{\infty}g(u)d(\exp(-e^{-u})). \tag{2.44}$$

In addition, one can show that the ratio of the standardizing function to the tail quantile or the translated tail quantile function converges to the EVI,

$$\lim_{x\to\infty}\frac{a(x)}{U(x)}\to 0,\ \lim_{x\to\infty}\frac{a(x)}{\inf\{x:F(x)=1\}-U(x)}\to 0, \tag{2.45}$$

when $\inf\{x:F(x)=1\}-U(x)$ is finite.

3. *if $\gamma<0$, one has the extremal-Weibull case characterized by*

$$\mathbb{E}\left[g\left(a_n^{-1}(X_{(n)}-b_n)\right)\right]\to\int_{-\infty}^{-\gamma^{-1}}g(u)d(\exp(-(1+\gamma u)^{-1/\gamma})). \tag{2.46}$$

In addition, one can show that the ratio of the standardizing function to the translated tail quantile function converges to the EVI,

$$\lim_{x\to\infty}\frac{a(x)}{\inf\{x:F(x)=1\}-U(x)}\to-\gamma, \tag{2.47}$$

and, furthermore, that $inf\{x:F(x)=1\}-U(x)$ is of the same regular variation as standardizing function $a(\cdot)$.

Furthermore, in Degen & Embrechts (2011, proposition 2.1) and Haan & Ferreira (2006, theorem 1.1.6), a unified domain of attraction condition combining these results for the complete range of $\gamma\in\mathbb{R}$ was provided and is detailed in Proposition 2.4.

Proposition 2.4 *For $\gamma\in\mathbb{R}$, the following are equivalent:*

1. *There exists a constant $a_n>0$ and $b_n\in\mathbb{R}$ such that*

$$\lim_{n\to\infty}F^n(a_nx+b_n)=H_\gamma(x)=\exp(-(1+\gamma x)^{-1/\gamma}), \tag{2.48}$$

for all $1+\gamma x>0$, which holds with $b_n=U(n)$ and $a_n=a(n)$.

2. *There exists a measurable function $a(\cdot)$ such that for $x>0$, the following holds for $U(y)=F^{\leftarrow}(1-1/y)$,*

$$\lim_{y\to\infty}\frac{U(yx)-U(y)}{a(y)}=\frac{x^\gamma-1}{\gamma}. \tag{2.49}$$

3. *There exists a measurable function* $f(\cdot) > 0$ *such that*

$$\lim_{y \to x_*} \frac{\overline{F}(y + xf(y))}{\overline{F}(y)} = (1 + \gamma x)^{-1/\gamma} \tag{2.50}$$

for all $1 + \gamma x > 0$*, which holds with* $f(y) = a(1/\overline{F}(y))$.

Now the conditions required for the existence and uniqueness of a limiting distribution for the standardized maximum loss have been derived; it is practically of interest to ask, given a severity distribution F that is continuous, can anyone say anymore about the extremal laws for each range of EVI considered. If possible, one would like these results to be re-expressed in terms of the distribution from which the losses were assumed realizations.

This leads one to redefine the results presented in Theorem 2.2 directly in terms of the severity distribution function of the OpRisk loss process. The result characterizing this re-expression of the 'extremal limit problem' in terms of the loss distribution is in Theoremnobreak 2.3 and is known as the celebrated *Fisher–Tippet–Gnedenko theorem*, which we present in the context of OpRisk modeling.

Theorem 2.3 (Fisher–Tippet–Gnedenko) *The limiting distributions (extremal laws) can be categorized according to the distribution for the maximal loss* $X_{(n)}$ *from a heavy-tailed severity model* $F(x)$ *which agrees in type with one of the following* $(\alpha > 0)$*:*

1. *The Frechet–Pareto case characterized by distribution*

$$\Phi_\alpha(x) = \exp(-x^{-\alpha}), \ \textit{if } x \geq 0, \tag{2.51}$$

and $\Phi_\alpha(x) = 0$ *otherwise. Furthermore, severity distribution* $F \in \mathscr{D}(\Phi_\alpha)$ *if and only if the following limit is satisfied:*

$$\lim_{x \to \infty} \frac{1 - F(\lambda x)}{1 - F(x)} \to \lambda^{-\alpha}, \ \textit{for all } \lambda > 0. \tag{2.52}$$

2. *The Gumbel case characterized by distribution*

$$\Lambda(x) = \exp(-e^{-x}), \ \textit{for } x \in \mathbb{R}. \tag{2.53}$$

Furthermore, severity distribution $F \in \mathscr{D}(\Lambda)$ *if and only if there exists an auxiliary function* $b(\cdot)$ *such that following limit is satisfied:*

$$\lim_{x \to \infty} \frac{1 - F(x + tb(x))}{1 - F(x)} \to \exp(-t), \ \textit{for all } t > 0. \tag{2.54}$$

3. *The Weibull case characterized by distribution*

$$\Psi_\alpha(x) = \exp(-|x|^\alpha), \ \textit{for } x \leq 0, \tag{2.55}$$

and $\Psi_\alpha(x) = 1$ *otherwise. Furthermore, severity distribution* $F \in \mathscr{D}(\Psi_\alpha)$ *if and only if, when* $\inf\{x : F(x) = 1\}$ *is finite, the following limit is satisfied:*

$$\lim_{x \to \infty} \frac{1 - F\left(\inf\{x : F(x) = 1\} - (\lambda x)^{-1}\right)}{1 - F\left(\inf\{x : F(x) = 1\} - x^{-1}\right)} \to \lambda^{-\alpha}, \ \textit{for all } \lambda > 0. \tag{2.56}$$

It is also practically useful to observe the following result from von Mises (1936), relating to the membership of a severity distribution in the maximal domain of attraction for the Frechet–Pareto class.

Theorem 2.4 (von Mises Condition) *Let F be a continuous distribution function which admits a density f satisfying the limit*

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \alpha > 0, \tag{2.57}$$

then $F \in \mathcal{D}(\Phi_\alpha)$, that is distribution F is in the domain of attraction of the Frechet-Pareto class.

Remark 2.3 *It is well known that the maximal domain of attraction of the Frechet–Pareto distribution consists purely of the distributions that satisfy the 'von Mises condition' and their tail-equivalent distributions.*

2.4 Generalized Extreme Value Loss Distributional Approach (GEV-LDA)

In this section, we consider the family of generalized extreme value loss distributional approach (GEV-LDA) models in the OpRisk context. We first complete the basic characterization of the three classes of extremal limit distributions specified by the Fisher–Tippet–Gnedenko theorem; we provide a set of examples illustrating which severity distributions will have maximal losses that have as their domain of attraction one of the three classes: Frechet–Pareto, Gumbel or Weibull. We select a few examples that are flexible in the properties of skewness and kurtosis they possess and can be of direct use to practitioners. We provide a brief introduction to the models presented, followed by their properties and a brief summary of how to perform estimation and Monte Carlo simulation for such models. Then we detail examples of the behaviour of these LDA models in several Poisson compound process examples that illustrate the effect on the annual loss estimated cdf versus changes in the Extreme Value Index (EVI) as well as the influence of the EVI on risk measures such as Value at Risk (VaR) under such models. These examples are intended to illustrate the importance of estimation and consideration of such heavy-tailed models in an OpRisk setting, to ensure that appropriate capital reserves are adequately modeled.

2.4.1 STATISTICAL CONSIDERATIONS FOR APPLICABILITY OF THE GEV MODEL

Having derived several properties of the distribution of the largest loss under a 'block maxima' approach, we note that the practical application of such an approach, including the estimation of the tail index, is reliant upon a few key statistical assumptions detailed later. In general, we note that under the block maxima approach to EVT, in an OpRisk setting, the risk modeller decides upon a suitable block period upon which to study the largest loss, for example, the largest monthly or annual loss. The scale or resolution of the modeling will clearly depend on the total number of losses observed in the given time scale.

Basic Statistical Assumptions for EVT-Block Maxima Approach in OpRisk.

- The width of each block of grouped losses, as measured, for example, in the OpRisk setting by weekly, monthly, quarterly or annual time periods is assumed to contain an equal number n of losses in each period (block).
- The number of losses per block n is assumed to be a known deterministic quantity. In practice, this is assumed to be large for the asymptotic results to be reasonable.
- The number of losses per block n is assumed constant and not time varying. In OpRisk, we know that we work with a compound process for each single-risk process; therefore, we will always encounter *a priori* an unknown random number of losses. So we refer to the average behaviour here and discuss extensions to tackle this later in the chapter.
- The n losses in each block are assumed to be independent both within a block and between blocks, and each loss is assumed to be identically distributed. This is not the typical approach in OpRisk modeling, because it is known that dependence between several risk processes occurs routinely in practice and varies potentially over time in a real financial operating environment. Again, we discuss the impact of non-i.i.d. losses in a subsequent section.

Having detailed these basic assumptions, we note that there have been several extensions to relax some of these features, and we will discuss these in the following sections. We also note that in practice, one does not need to start with a parametric assumption on the severity model, to arrive at a distribution for the maximal loss in the GEV family, instead it is common to start with the GEV family and to estimate the parameters of this model, either under restrictions on the parameter space in a frequentist setting or with *a priori* beliefs about appropriate signs or ranges for the EVI. To achieve the fitting in this context, it is required to specify the likelihood, on the time resolution on which we consider the maximum, for example, weekly, monthly or yearly. In Equation 2.58, we present the log-likelihood for severity models constructed utilizing the GEV distribution. In working with this likelihood, the data is prepared as detailed below.

Algorithm 2.1 (Preparing the Data under a Block Maxima Approach)

1. *Take the full set of N losses observed in a period $[0, T]$, time ordered and denoted by $\{X_1(\tau_1), \ldots, X_N(\tau_N)\}$ for $0 \leq \tau_1 < \tau_2 < \cdots < \tau_N \leq T$, where the ith loss occurred at time τ_i. Then portion them into the resolution desired, weekly, monthly, quarterly, yearly, etc. to obtain sets of n (approximately n and required to be large in practice) losses for each unit of time block indexed by k. This produces sets of losses denoted for the kth block of time unit by $\{\widetilde{X}_1(k), \ldots, \widetilde{X}_n(k)\}$, where $\widetilde{X}_i(k)$ denotes the i-th loss in block k. This is repeated for all blocks $k \in \{1, 2, \ldots, K\}$.*
2. *Sort the data in each block to obtain the order statistics, where for the k-th block one obtains $\{\widetilde{X}_{(1)}(k), \ldots, \widetilde{X}_{(n)}(k)\}$.*
3. *Then, for example, record the largest loss per block to obtain the set*

$$\boldsymbol{X}_{(n)} = \{\widetilde{X}_{(n)}(1), \ldots, \widetilde{X}_{(n)}(K)\}$$

of maximum losses over the K blocks. This is the data utilized under the block maxima approach to fit the likelihood model.

Then under the assumptions detailed earlier, the likelihood for the GEV as a function of this maximum loss per block of data is given for $\gamma \neq 0$ by

$$\begin{aligned}\ell(\widetilde{\mathbf{X}}_{(n)};\mu,\sigma,\gamma) &= \ln\left[\prod_{k=1}^{K} H_\gamma\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right)\right] \\ &= -K\ln(\sigma) - (1+1/\gamma)\sum_{k=1}^{K}\ln\left[1+\gamma\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right)\right] \\ &\quad -\sum_{k=1}^{K}\left[1+\gamma\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right)\right]^{-1/\gamma} \end{aligned} \tag{2.58}$$

with the condition,

$$1+\gamma\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right) > 0, \tag{2.59}$$

which ensures that the log-likelihood is finite. If this condition is violated by certain parameter values, then one of the observed maximum falls beyond the end point of the distribution.

If $\gamma = 0$, the likelihood is given according to the Gumbel limit of the GEV distribution,

$$\begin{aligned}\ell(\widetilde{\mathbf{X}}_{(n)};\mu,\sigma,0) &= \ln\prod_{k=1}^{K} H_0\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right) \\ &= -K\ln(\sigma) - \sum_{k=1}^{K}\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right) - \sum_{k=1}^{K}\exp\left[-\left(\frac{\widetilde{X}_{(n)}(k)-\mu}{\sigma}\right)\right]. \end{aligned} \tag{2.60}$$

Next we present a brief discussion on the different approaches statisticians have adopted for estimation of the GEV model parameters.

2.4.2 VARIOUS STATISTICAL ESTIMATION PROCEDURES FOR THE GEV MODEL PARAMETERS IN OpRisk SETTINGS

Given the likelihood, the moments of the GEV distribution or the quantile function, there are numerous statistical approaches one could adopt to perform the parameter estimation. These could include the approach of maximum-likelihood estimation (MLE) Prescott and Walden (1980); Hosking *et al.* (1985); Macleod (1989), methods of moments and the generalization of this approach such as Hosking (1990) or quantile-based matching approaches. In the case of the MLE, it was shown in Smith (1990) that estimators under this approach are asymptotically efficient when $\gamma < 0.5$ and when the generalized method of L-moments is considered, the estimates of the parameters are asymptotically consistent, with Gaussian distribution when $\gamma > -0.5$.

Remark 2.4 *In general, it is reported in the literature that when small sample sizes K are considered, such as would be typical in an OpRisk setting, the approach of method of moments (MOM) is advocated over MLE because, in general, the accuracy is greater. However, if the sample size is small and, in addition, constraints are imposed on the parameter space over which solutions are considered, then in such cases one is recommended to resort back to the MLE approach for improved accuracy. In addition, under strict equality constraints, one should resort to profile-likelihood-based*

methods. Finally, we note that in small sample cases, it is also strongly recommended to consider utilizing Bayesian approaches such as those of Martins and Stedinger (2000), Coles and Powell (1996), Eugenia and Cabras (2007) or Lopes et al. (2011) which allow the modeller to incorporate a priori knowledge of the possible model parameter ranges into the inference.

So a likelihood-based approach to the parameter estimation typically involves finding the arguments that maximize the likelihood for the observed realizations of maximum losses in the specified blocks. This typically involves finding the partial derivatives of the GEV likelihood with respect to the parameters, setting them to zero and then using a Newton–Raphson iteration procedure to solve for the parameters as in Prescott & Walden (1980).

Remark 2.5 *The solution to the MLE is non-analytic. Therefore, it must be solved numerically and can be achieved via a standard numerical optimization package utilizing gradient descent, simplex methods or the Newton–Raphson approach. However, the following cautions must be considered: ensure that when $\gamma \neq 0$, the constraint in Equation 2.59 is strictly adhered to by the numerical optimization package. In addition, when $\gamma \in [-\varepsilon, \varepsilon]$ for some small ε, then one should switch to the likelihood model in Equation 2.60 to avoid numerical difficulties. Finally, it is worth noting that the popular Newton–Raphson approaches can only provide local convergence and may fail when the objective function becomes non-convex. Some authors have, therefore, advocated the use of gradient descent without the use of second-order information from the Hessian, arguing that complications due to non-convexity may be avoided in such an approach at a loss of computational efficiency.*

Alternatively, one may also consider a Bayesian approach, under which one would specify priors on the parameters and combine this likelihood model into a posterior inference setting, with sampling or maximization or both possible. Such approaches are detailed for the GEV model in Coles and Powell (1996) and Shevchenko (2011).

Next we discuss the proposed mixed-type estimation approaches in the literature, which aim to utilize combinations of the knowledge of the parametric forms of the moments and the likelihood to capture the best statistical features of both methods.

2.4.2.1 Mixed MLE and Method of L-Moments and L-Trimmed-Moments-Based Approaches.

The mixed MLE and L-moments-based approach of Morrison and Smith (2002) involves combining the knowledge of the moments of the GEV distribution with respect to the parameters, with the method of MLE. There are several approaches that can be considered in this regard; however, all approaches aim to incorporate the individual statistical features of efficiency, consistency and unbiasedness of the moment and MLE-based approaches to combine these properties of each approach to tackle scenarios with small to moderate data sizes, where the asymptotic assumptions may not be suitable to justify distributional properties of the estimators.

Each of these mixed-type approaches effectively involves the specification of a set of one-to-one correspondences or functions between a subset of the GEV model parameters and new parameters and the remainder of the GEV model parameters. This is typically done in a setting in which one has access to known unbiased estimators defined with respect to the maximum losses in each block for the new parameters under the mapping. Then the remaining model parameters from the original GEV specification are estimated via MLE methods after substitution of the estimators for the new model parameter estimates is performed. This could be, for example, mappings from the model parameters to moment conditions and MLE of the EVI parameter γ.

Generically, this idea can be written down according to the following relationship in which one defines for the original GEV model parameters $\boldsymbol{\theta} = (\mu, \sigma, \gamma)$ a new set of parameters

$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ via a one-to-one mapping. Many choices of mapping are possible and will depend on the modellers aims and should, in general, take into consideration additional features such as the ability to utilize data to estimate some of the new model parameters in an unbiased manner. That is, if there is knowledge of unbiased estimators for some of these parameters, say, for example, $\widehat{\lambda}_1 = \widehat{\lambda}_1(\widetilde{\boldsymbol{X}}_{(n)})$ and $\widehat{\lambda}_2 = \widehat{\lambda}_2(\widetilde{\boldsymbol{X}}_{(n)})$, then the estimate of the remaining parameter can be performed via MLE on the reparameterized model according to

$$\widehat{\lambda}_3 = \arg\max_{\lambda_3} \ell(\widehat{\lambda}_1, \widehat{\lambda}_2, \lambda_3). \tag{2.61}$$

Then to recover the original estimates of the GEV model parameters, one has under the one-to-one mapping the inverse solution given by $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\widehat{\boldsymbol{\lambda}})$.

One of the more utilized forms of mapping involves the L-moments of a distribution with respect to the parameters. The L-moments, given by Hosking (1990), are defined for real-valued random variable X with distribution $F(x)$ and quantile function $Q(p)$ according to Definition 2.17. Effectively, the L-Moments correspond to linear combinations of the order statistics and, in addition, have the property that compared to conventional moments; they have lower sample variances and are more robust against outliers.

The development of L-Moments-based ideas go back to the study of linear combinations of order statistics (Sillitto, 1969; David, 1968; Chernoff *et al.*, 1967; Greenwood *et al.*, 1979); then Hosking (1990) combined these results to complete a specification of the representation of a distribution with respect to its L-moments uniquely. In addition to the theory of L-moments, there have also been developments of LQ-moments by Mudholkar and Hutson (1998) and trimmed versions TL-moments by Elamir and Seheult (2003) and given in Definition 2.17, Definition 2.19 and Definition 2.20.

Definition 2.17 (Distributional L-Moments) *The L-moments of a real-valued random variable $X \sim F(x)$ with quantile function $Q(u) = F^{-1}(u)$, for which there is an r sample realization with order statistics given by $X_{(1)}(r) \leq X_{(2)}(r) \leq \ldots \leq X_{(n)}(r) \leq \ldots \leq X_{(r)}(r)$, is defined for each value of r according to either linear combinations of the expectations of the order statistics or alternatively via the Fourier coefficients of the quantile function in terms of orthogonal polynomials on the interval $[0, 1]$ producing the two following representations*

$$\begin{aligned}\lambda_r &= \int_0^1 Q(F) P^*_{r-1}(F)\, dF \\ &= r^{-1} \sum_{k=0}^{r-1} (-1)^k C_k^{r-1} \mathbb{E}[X_{(r-k)}(r)],\ \forall r = 1, 2, \ldots,\end{aligned} \tag{2.62}$$

with shifted Legendre polynomials $P^*_r(F) = \sum_{k=0}^r p^*_{r,k} F^k$ and $p^*_{r,k} = (-1)^{r-k} C_k^r C_k^{r+k}$ and where C_k^r represents r choose k which is given by

$$C_k^r := \frac{r!}{(r-k)!k!} \tag{2.63}$$

Equivalently, one can consider the expectation of the n-th order statistic from a sample of size r as given by (David & Nagaraja 1970, p.33)

$$\begin{aligned}\mathbb{E}[X_{(n)}(r)] &= \frac{r!}{(n-1)!(r-n)!} \int Q(F(x))^{n-1} (1 - F(x))^{r-n}\, dF(x) \\ &= \frac{r!}{(n-1)!(r-n)!} \int_0^1 Q(u) u^{n-1} (1-u)^{r-n}\, du.\end{aligned} \tag{2.64}$$

Here λ_r is a linear function of the expected order statistics.

Remark 2.6 *As realized in Perez et al. (2003), these specifications of a distribution and its properties through the L-moments, therefore, falls into a class of L-functionals specified according to the integral operation on the quantile function given generically by Equation 2.65, where the notation* $\lambda_r = L_r(F)$ *is adopted to show that the* r*-th L-moment is directly an integral functional of the quantile function* $Q(u)$ *corresponding to the distribution function* F*,*

$$\lambda_r = \int_0^1 Q(u)J(u)\,du \tag{2.65}$$

for some $J(u)$*, which is a bounded and measurable function on* $[0, 1]$*. Consequently, one can show that the estimation of the L-moments based on sample empirical distributions* F_n *can satisfy a CLT result,*

$$\sqrt{n}\,(L_r(F) - L_r(F_n)) \to Normal\left(0, \sigma^2(F)\right) \tag{2.66}$$

with

$$\sigma^2(F) = \int_0^2 \int_0^1 J(u)J(v)u(1-v)Q(u)Q(v)\,dvdu, \tag{2.67}$$

see Serfling (1981).

For more specific properties of robustness, global efficiency of the L-moment estimators, the interested reader is referred to Perez *et al.* (2003).

Remark 2.7 *It turns out that a consequence of this definition of L-moments is that one can easily find estimators of these quantities based on an observed sample of data, simply by using linear combinations of the ordered data values, that is, an L-statistic.*

Therefore, one can express the first few L-moments according to the following identities in Definition 2.18 for a sample of size K.

Definition 2.18 *The first four L-moments of a real-valued random variable* $X \sim F(x)$*, for which there is a* K *sample realization with order statistics given by* $X_{(1)}(K) \leq X_{(2)}(K) \leq \cdots \leq X_{(n)}(K) \leq \cdots \leq X_{(K)}(K)$*, are given according to*

$$\begin{aligned}
\lambda_1 &= \mathbb{E}[X_{(1)}(1)] = \int_0^1 Q(u)\,du, \\
\lambda_2 &= \frac{1}{2}\mathbb{E}\left[X_{(2)}(2) - X_{(1)}(2)\right] = \int_0^1 Q(u)(2u-1)\,du, \\
\lambda_3 &= \frac{1}{3}\mathbb{E}\left[X_{(3)}(3) - 2X_{(2)}(3) + X_{(1)}(3)\right] = \int_0^1 Q(u)(6u^2 - 6u + 1)\,du, \\
\lambda_4 &= \frac{1}{4}\mathbb{E}\left[X_{(4)}(4) - 3X_{(3)}(4) + 3X_{(2)}(4) - X_{(1)}(4)\right] \\
&= \int_0^1 Q(u)(20u^3 - 30u^2 + 12u - 1)\,du.
\end{aligned}$$

Analogously to conventional moments, the L-moments of orders one to four characterize location, scale, skewness and kurtosis, respectively.

The sample estimators of the first four L-moments are provided after Definition 2.22 by Equation 2.86 and Equation 2.87. In addition, according to Theorem 2.5 of Hosking (1990), the L-moments can be used to uniquely and exactly characterize a distribution.

Theorem 2.5 (L-Moment Representation) *The L-moments λ_r for all $r = 1, 2, \ldots,$ of a real-valued random variable X exist if and only if the distribution for X has finite mean. In situations in which this mean is finite, then the distribution is uniquely characterized by its set of L-moments $\{\lambda_r : r \in \mathbb{J}^+\}$.*

The robustification of the L-moments was achieved by the definition of the LQ-moments, which replace the mean of the distribution of the random variable $X_{(r-k)}(r)$ with a robust version such as the median or a different location measure. The LQ-moments are given by the generic representation in Definition 2.19.

Definition 2.19 (Distributional LQ-Moments) *The LQ-moments of a real-valued random variable $X \sim F(x)$ with a quantile function $Q_X(u) = F_X^{-1}(u)$, for which there is an r sample realization with order statistics given by $X_{(1)}(r) \leq X_{(2)}(r) \leq \cdots \leq X_{(n)}(r) \leq \cdots \leq X_{(r)}(r)$, is defined for each r according to*

$$\zeta_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k C_k^{r-1} \tau_{p,\alpha}[X_{(r-k)}(r)], \ \forall r = 1, 2, \ldots, \tag{2.68}$$

where $\tau_{p,\alpha}[X_{(r-k)}(r)]$ is given by

$$\tau_{p,\alpha}[X_{(r-k)}(r)] = pQ_{X_{(r-k)}(r)}(\alpha) + (1-2p)Q_{X_{(r-k)}(r)}\left(\frac{1}{2}\right) + (1-2p)Q_{X_{(r-k)}(r)}(1-\alpha) \tag{2.69}$$

with $0 \leq \alpha \leq \frac{1}{2}$ and $0 \leq p \leq \frac{1}{2}$.

Examples of proposed choices of values of (α, p) that correspond to standard location estimations are given in the following remark.

Remark 2.8 *The choice of $\tau_{p,\alpha}[X_{(r-k)}(r)]$ can be selected to make the estimation of the LQ-moments from samples simply through the use of one of the following choices:*

1. *$\tau_{\alpha,p}$ corresponds to the median for $(\alpha = 1, p = 0)$;*
2. *$\tau_{\alpha,p}$ corresponds to the mean for $(\alpha = \frac{1}{4}, p = \frac{1}{4})$;*
3. *$\tau_{\alpha,p}$ corresponds to Gastwirth weighted median estimator for $(\alpha = \frac{1}{3}, p = 0.3)$.*

Perhaps a simpler specification in certain settings is to consider the trimmed L-moments, that is, the TL-moments, which are specified to be robust estimators, assign zero weight to extreme outliers and, in addition, are always defined for a population even in cases in which L-moments are not defined for a population from which a sample of observed losses is realized. For example, if the severity distribution is Cauchy, in which the standard first moment does not exist and the L-moments do not exist, however, the population TL-moment is well defined.

Definition 2.20 (Distributional Trimmed L-Moments) *Consider a real-valued random variable $X \sim F(x)$. Given a sample size of r with an actual sample size of $r + t_1 + t_2$ with*

order statistics given by $X_{(1)}(r+t_1+t_2) \leq X_{(2)}(r+t_1+t_2) \leq \cdots \leq X_{(n)}(r) \leq \cdots \leq X_{(r+t_1+t_2)}(r+t_1+t_2)$*, consider only the selection of the* r *order statistics in the set given by* $X_{(t_1+1)}(r+t_1+t_2) \cdots \leq X_{(t_1+r)}(r)$*, obtained by trimming the* t_1 *smallest and the* t_2 *largest from the sample. Then the TL-moments are defined for each combination* (r, t_1, t_2) *by*

$$\lambda_r^{(t_1,t_2)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k C_k^{r-1} \mathbb{E}\left[X_{(r+t_1-k)}(r+t_1+t_2)\right], \ \forall r = 1, 2, \ldots, \tag{2.70}$$

with the moments given by

$$\begin{aligned} \mathbb{E}[X_{(n)}(r)] &= \frac{r!}{(n-1)!(r-n)!} \int Q(F(x))^{n-1}(1-F(x))^{r-n} dF(x) \\ &= \frac{r!}{(n-1)!(r-n)!} \int_0^1 Q(u) u^{n-1}(1-u)^{r-n} du. \end{aligned} \tag{2.71}$$

In general, we can also note the following relationship between the L-moments and the TL-moments.

Remark 2.9 *It is clear that the TL-moments encompass the L-moments in the setting in which* $t_1 = t_2 = 0$*, and in general, the amount of trimming from above or below does not have to be equal, i.e.* $t_1 \neq t_2$*. In addition, it can be noted that in the symmetric case in which* $t_1 = t_2 = t$ *the TL-moment* $\lambda_r^{(t)}$ *corresponds to consideration of the expectation of the median from a conceptual sample of size* $1 + 2t$ *and is, therefore, zero for symmetric distributions.*

Finally, it will be of benefit in utilizing the L-moments in the estimation of the GEV model parameters to also define the L-skewness and L-kurtosis according to the Definition 2.21.

Definition 2.21 (Distributional L-Skewness, L-Kurtosis) *The L-moment definition of skewness and kurtosis are given, respectively, by*

$$\tau_3 = \frac{\lambda_3}{\lambda_2}, \ \tau_4 = \frac{\lambda_4}{\lambda_2}. \tag{2.72}$$

The equivalent LQ-moment definitions are given by

$$\eta_3 = \frac{\zeta_3}{\zeta_2}, \ \eta_4 = \frac{\zeta_4}{\zeta_2}, \tag{2.73}$$

and, finally, the TL-moment definitions are given by

$$\tau_3^{t_1,t_2} = \frac{\lambda_3^{t_1,t_2}}{\lambda_2^{t_1,t_2}}, \ \tau_4^{t_1,t_2} = \frac{\lambda_4^{t_1,t_2}}{\lambda_2^{t_1,t_2}}. \tag{2.74}$$

Having defined the L-moments for linear combinations of order statistics in terms of the population (severity distribution), then it is practically useful to consider the case in which, given a realization of samples from this population, the estimators of these moments can be obtained from the sample L-moments. Typically, the estimators of the L-moments obtained

from a particular sample (efficient for normal populations) are given by those derived in Downton (1966) and shown in Lemma 2.1 for the TL-moments and in Equation 2.86 for sample estimators of L-moments.

Lemma 2.1 (Sample Estimators of TL-Moments) *Given a sample of size n with order statistics $X_{(1)}(n) \leq X_{(2)}(n) \leq \cdots \leq X_{(n)}(n)$. Consider the linear combinations of order statistics defined by*

$$U_{(kl)}(n) = k!l! \sum_{i=1}^{n} C_k^{i-1} C_l^{n-i} X_{(i)}(n), \tag{2.75}$$

for which the following relationship holds,

$$\mathbb{E}[U_{(kl)}(n)] = k!l! C_{k+l+1}^n \mathbb{E}\left[X_{(k+1)}(k+l+1)\right]. \tag{2.76}$$

It was shown in Downton (1966) that an unbiased estimator of the expectation of the $(k+1)$-th order statistic $X_{(k+1)}(k+l+1)$ in a conceptual sample of size $k+l+1$ can be obtained according to the following sample estimator of the expectation,

$$\widehat{\mathbb{E}}\left[X_{(k+1)}(k+l+1)\right] = \left(C_{k+l+1}^n\right)^{-1} \sum_{i=1}^{n} C_k^{i-1} C_l^{n-i} X_{(i)}(n). \tag{2.77}$$

The equivalent unbiased estimator that would be used to replace the expectation in the r-th TL-moments $\lambda_r^{t,t}$ is given for $t_1 = t_2 = t$ by the following sample estimator of the expectation,

$$\widehat{\mathbb{E}}\left[X_{(r+t-k)}(r+2t)\right] = \left(C_{r+2t}^n\right)^{-1} \sum_{i=1}^{n} C_{r+t-k-1}^{i-1} C_{t+k}^{n-i} X_{(i)}(n). \tag{2.78}$$

To complete the estimation of the population L-moments and TL-moments with the sample estimators, we simply take the linear combinations given by

$$\widehat{\lambda}_r^{(t,t)} = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k C_k^{r-1} \widehat{\mathbb{E}}\left[X_{(r+t-k)}(r+2t)\right]. \tag{2.79}$$

Note for the sample estimates of the standard L-moments, simply set $t = 0$.

For completeness, we also note that exact expressions for variance of TL-moment estimators are derived analytically in Elamir & Seheult (2003) and presented in Lemma 2.2.

Lemma 2.2 (Variance of TL-Moment Estimators) *Analytic expressions for the variance of the TL-moment estimators can also be obtained to aid in the estimation of the precision of these quantities for an actual sample of size n. In the context of this section, the estimation accuracy of the GEV model parameters is given according to*

$$\mathbb{E}[\widehat{\lambda}_r^{(t,t)} \widehat{\lambda}_s^{(t,t)}]$$

$$= \frac{1}{rs} \sum_{k=0}^{r-1} \sum_{l=0}^{s-1} (-1)^{r+s} C_k^{r-1} C_l^{s-1} \Omega(r+t-k-1, k+t, s+t-l-1, l+t), \tag{2.80}$$

where

$$\Omega(k,l,p,q) = w_0(k,l,p,q)\left[\sum_{r=0}^{p}\sum_{s=0}^{q} w_1(k,l,p,q,r,s)\,\mathbb{E}\left[X^2_{(k+r+1)}(k+l+r+s+1)\right]\right.$$

$$+\sum_{r=0}^{l}\sum_{s=0}^{p} w_2(k,l,p,q,r,s)\,\mathbb{E}\left[X_{(k+r+1)}(k+q+r+s+2)X_{(k+r+2)}(k+q+r+s+2)\right]$$

$$\left.+\sum_{r=0}^{q}\sum_{s=0}^{k} w_3(k,l,p,q,r,s)\,\mathbb{E}\left[X_{(p+r+1)}(p+l+r+s+2)X_{(p+r+2)}(p+l+r+s+2)\right]\right]$$

with

$$\begin{aligned}
w_0(k,l,p,q) &= (k!l!p!q!C^n_{k+l+1}C^n_{p+q+1})^{-1}\\
w_1(k,l,p,q,r,s) &= (p-r)!(q-s)!(k+r)!(l+s)!C^p_rC^q_sC^k_{p-r}C^l_{q-s}C^n_{k+l+r+s+1}\\
w_2(k,l,p,q,r,s) &= (-1)^{r+s}(l-r)!(p-s)!(k+r)!(q+s)!C^l_rC^p_SC^{n-k-r-1}_{l-r}\\
&\quad\times C^{n-q-s-1}_{p-s}C^n_{k+q+r+s+2}\\
w_3(k,l,p,q,r,s) &= (-1)^{r+s}(q-r)!(q-s)!(p+r)!(l+s)!C^q_rC^k_sC^{n-p-r-1}_{q-r}\\
&\quad\times C^{n-l-s-1}_{k-s}C^n_{p+l+r+s+2}.
\end{aligned}$$

In addition, distribution-free unbiased estimators of the variances and covariances of sample TL-moments can be obtained to approximate these expressions utilizing the identities given by

$$\widehat{\mathbb{E}}[X_{(k+1)}(k+l+1)] = \left(C^n_{k+l+1}\right)^{-1}\sum_{i=1}^{n} C^{i-1}_k C^{n-i}_l X_{(i)}(n)$$

$$\widehat{\mathbb{E}}[X^2_{(k+1)}(k+l+1)] = \left(C^n_{k+l+1}\right)^{-1}\sum_{i=1}^{n} C^{i-1}_k C^{n-i}_l X^n_{(i)}(n)$$

$$\widehat{\mathbb{E}}[X_{(k+1)}(k+l+2)X_{(k+2)}(k+l+2)] = \left(C^n_{k+l+2}\right)^{-1}\sum_{i=1}^{n}\sum_{j=i+1}^{n} C^{i-1}_k C^{n-i}_l X_{(i)}(n)X_{(j)}(n).$$

to substitute into the expressions for the moments and mixed moments.

One may now proceed to define a mixed strategy for parameter estimation under the GEV model. The first of these approaches we detail was proposed explicitly by Morrison & Smith (2002), and it adopts the following stages. First, it requires the one-to-one mappings to be constructed from knowledge of the L-method of moment (LMOM) results given for each of the parameters of the GEV distribution in Definition 2.22 due to Hosking (1990).

Definition 2.22 (L-Method of Moments for GEV) *The population LMOM expressions for the parameters of the GEV distribution are given by solutions, with respect to the L-skewness, by the system of equations*

$$\frac{1-3^{\widehat{\gamma}}}{1-2^{\widehat{\gamma}}} = \frac{\widehat{\tau}_3+3}{2},$$

$$\widehat{\sigma} = -\frac{\widehat{\lambda}_2\widehat{\gamma}}{(1-2^{\widehat{\gamma}})\Gamma(1-\widehat{\gamma})},$$
$$\widehat{\mu} = \widehat{\lambda}_1 + \frac{\widehat{\sigma}}{\widehat{\gamma}}[1-\Gamma(1-\widehat{\gamma})],$$

where $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ *and* $\widehat{\tau}_3 = \widehat{\lambda}_3/\widehat{\lambda}_2$ *are the estimators of the first two LMOMs and the L-skewness obtained from the sample.*

Obtaining these L-moment estimators $\widehat{\lambda}_1, \widehat{\lambda}_2, \widehat{\lambda}_3$ and the L-skewness $\widehat{\tau}_3$ involves an unbiased estimator of the first three probability-weighted moments given by Equation 2.81 in which $\lambda_1 = \beta_0$, $\lambda_2 = 2\beta_1 - \beta_0$ and $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0$, where

$$\beta_r = \mu - \frac{\sigma}{\gamma}\frac{[1-(r+1)^{\gamma}\Gamma(1-\gamma)]}{r+1}. \tag{2.81}$$

In Landwehr *et al.* (1979), an unbiased estimator of these probability-weighted moments (PWM) in terms of the order statistics, for a sample of size n, is provided according to Equation 2.82

$$\widehat{\beta}_r = \sum_{i=1}^{n}\left[\frac{(i-1)(i-2)\cdots(i-r)}{n(n-1)\cdots(n-r)}x_{(i)}(n)\right], \quad \forall r \in \{0,1,2,\ldots,n\}, \tag{2.82}$$

where $x_{(i)}(n)$ is the realized *i*-th order statistic of a sample of size *n*. The following properties of these estimators is discussed in Morrison & Smith (2002) and detailed in Remark 2.10.

Remark 2.10 *The bias of the LMOM estimates of the shape parameter of the GEV distribution* γ *increases with a decrease in the magnitude of* γ *and can be significant. In contrast, the MLE method produces an almost unbiased estimate of* γ *for small samples; however, the variance of the MLE estimates is larger than those obtained from the biased LMOM approach. In addition, it has been reported in Martins and Stedinger (2000) that the MLE estimates of the shape parameter can regularly in practice produce unrealistic estimates of* γ*, resulting in very large errors in the quantile estimates. The mixed-type methods aim to overcome these difficulties.*

In practice, for such a mixed approach, the parameter range of the extreme value index is restricted to $\gamma \in [-0.5, 0.5]$ to ensure that the moments are finite and appropriate regularity conditions for the MLE are satisfied, see discussion on such items in the GPD setting and the GEV setting in Beirlant *et al.* (2004, chapter 5). However, this can be relaxed through the consideration of the TL-moments approach. The mixed MLE and L-moments-based approach is presented as follows.

Mixed MLE and Moment-Based Parameter Estimation in GEV Models

Stage 1. Reparameterize the GEV model likelihood in terms of only the EVI γ. This is achieved by expressing the parameters μ and σ as functions of γ via constraints on the L-moments given by

$$\lambda_1 = \mathbb{E}[X_{(1)}(1)] = \mu - \frac{\sigma}{\gamma}(1-\Gamma(1-\gamma))$$
$$\lambda_2 = \mathbb{E}[X_{(2)}(2) - X_{(1)}(2)] = -2\frac{\sigma}{\gamma}(1-2^{\gamma})\Gamma(1-\gamma). \tag{2.83}$$

Stage 2. Estimate the L-moments empirically utilizing the statistics given by

$$\begin{aligned}\widehat{\lambda}_1 &= \frac{1}{K}\sum_{k=1}^{K}\widetilde{X}_{(n)}(k)\\ \widehat{\lambda}_2 &= \frac{1}{K(K-1)}\sum_{i<j}^{K}|\widetilde{X}_{(n)}(i)-\widetilde{X}_{(n)}(j)|.\end{aligned} \tag{2.84}$$

Then utilize these estimates to obtain the functions for μ and σ with respect to γ according to the expressions incorporating these L-moment estimates given by

$$\begin{aligned}\sigma &= -\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\Gamma(1-\gamma)]^{-1}\\ \mu &= \widehat{\lambda}_1+\frac{1}{\gamma}(1-\Gamma(1-\gamma))\left\{-\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\,\Gamma(1-\gamma)]^{-1}\right\}.\end{aligned} \tag{2.85}$$

Stage 3. Perform MLE for the EVI parameter γ subject to the constraints on L-moments imposed by the estimates in Stage 2, which involves maximization of the reparameterized likelihood given, for $\gamma \neq 0$ by

$$\begin{aligned}\ell(\widetilde{\boldsymbol{X}}_{(n)};\mu,\sigma,\gamma) = &-K\ln\left(-\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\,\Gamma(1-\gamma)]^{-1}\right)\\ &-(1+1/\gamma)\sum_{k=1}^{K}\ln[1+\gamma\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma)]\\ &-\sum_{k=1}^{K}[1+\gamma\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma)]^{-1/\gamma},\end{aligned} \tag{2.86}$$

where the function $\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma)$ is defined according to

$$\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma) = \left(\frac{\widetilde{X}_{(n)}(k)-\left(\widehat{\lambda}_1+\frac{1}{\gamma}\,(1-\Gamma(1-\gamma))\left\{-\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\,\Gamma(1-\gamma)]^{-1}\right\}\right)}{-\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\Gamma(1-\gamma)]^{-1}}\right). \tag{2.87}$$

If $\gamma = 0$, the likelihood is given according to the Gumbel limit of the GEV distribution according to

$$\begin{aligned}\ell(\widetilde{\boldsymbol{X}}_{(n)};\mu,\sigma,0) = &-K\ln\left(-\frac{1}{2}\gamma\widehat{\lambda}_2[(1-2^{\gamma})\,\Gamma(1-\gamma)]^{-1}\right)\\ &-\sum_{k=1}^{K}\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma)-\sum_{k=1}^{K}\exp[-\mathcal{S}(\widetilde{X}_{(n)}(k),\gamma)].\end{aligned} \tag{2.88}$$

In addition, the properties of estimators obtained under this mixed method can be studied. In this context, if one can prove the model parameter estimates after inversion $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\widehat{\boldsymbol{\lambda}})$ are

consistent, then perhaps under certain regularity conditions on the GEV likelihood under the $\boldsymbol{\theta}$ parameterizations, such as restriction of $\gamma \in [-0.5, 0.5]$, then there will be a CLT result asymptotically with $K \to \infty$ involving

$$\sqrt{K}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \sim \mathrm{N}(\mathbf{0}, \mathbf{J}^T \Sigma \mathbf{J}) \tag{2.89}$$

in which the Jacobian matrix $\mathbf{J}$ is given by

$$\mathbf{J} = \frac{\partial \boldsymbol{\theta}^T}{\partial \boldsymbol{\lambda}} \tag{2.90}$$

and the matrix $\Sigma = \mathbf{U}\mathbf{V}\mathbf{U}^T$ is defined according to matrices

$$\mathbf{V} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & 0 \\ \Lambda_{21} & \Lambda_{22} & 0 \\ 0 & 0 & I^{\lambda}_{33} \end{bmatrix}, \ \mathbf{U}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ I^{\lambda}_{31} & I^{\lambda}_{32} & I^{\lambda}_{33} \end{bmatrix}, \tag{2.91}$$

where

$$\mathbf{I}^{\lambda} = \mathbf{J}\mathbf{I}^{\theta}\mathbf{J}^T, \ V_{ij} = \lim_{n\to\infty} n\mathbb{C}\mathrm{ov}\,(\widehat{\lambda}_i, \widehat{\lambda}_j), \tag{2.92}$$

with $\mathbf{I}^{\theta}$ denoting the information matrix of the GEV distribution. This is known and detailed in Equation 2.93 as provide in Beirlant *et al.* (2004, Appendix 5.9, p.169) according to

$$\mathbf{I}^{\theta} = -\mathbb{E}\left[\frac{\partial^2 \ell(x; \mu, \sigma, \gamma)}{\partial\, \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^T}\right] \tag{2.93}$$

with generic elements given by

$$I^{\boldsymbol{\theta}}_{11} = \frac{1}{\sigma^2\gamma^2}(1 - 2\Gamma(2+\gamma)),$$

$$I^{\boldsymbol{\theta}}_{12} = -\frac{1}{\sigma\gamma^2}\left(1 - e_C - q + \frac{1-\Gamma(2+\gamma)}{\gamma} + \frac{p}{\gamma}\right),$$

$$I^{\boldsymbol{\theta}}_{13} = -\frac{1}{\sigma^2\gamma}(p - \Gamma(2+\gamma)),$$

$$I^{\boldsymbol{\theta}}_{22} = \frac{1}{\gamma^2}\left[\frac{\pi^2}{6} + \left(1 - \gamma^* + \frac{1}{\gamma}\right)^2 - \frac{2q}{\gamma} + \frac{p}{\gamma^2}\right],$$

$$I^{\boldsymbol{\theta}}_{23} = -\frac{1}{\sigma\gamma}\left(q - \frac{p}{\gamma}\right),$$

$$I^{\boldsymbol{\theta}}_{33} = -\frac{p}{\sigma^2},$$

where e_C is Euler's constant and

$$\begin{aligned} p &= (1+\gamma)^2\Gamma(1+2\gamma), \\ g &= \Gamma(2+\gamma)\left(\Psi(1+\gamma)+\frac{1+\gamma}{\gamma}\right), \\ \Psi(x) &= \frac{d\ln\Gamma(x)}{dx}. \end{aligned} \tag{2.94}$$

Remark 2.11 *Under this mixed approach, the consistency of the estimator for $\widehat{\boldsymbol{\theta}}$ follows from Smith (1986). In addition, analytic expressions for the components of matrix V_{ij} can be obtained from standard theory of L-moments, see Hosking (1990) and U-statistics, see Lee (1990).*

2.4.2.2 Profile Likelihood and Confidence Intervals on the Return Level. The class of mixed procedures described earlier are examples of formulating a profile likelihood estimation procedure. Another, practically relevant problem in this class is to obtain estimates of the confidence intervals for the return levels of the GEV model.

The formal definition of the profile log-likelihood is provided in Definition 2.23 for a parameter block $\boldsymbol{\theta}_i$ such that the parameter vector is partitioned into any desired sub-blocks $\boldsymbol{\theta} = (\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i})$.

Definition 2.23 *The profile log-likelihood for parameter vector $\boldsymbol{\theta}_i$, which is a sub-block of the parameter vector $\boldsymbol{\theta}$, is given by*

$$\ell_p(\boldsymbol{\theta}_i) = \max_{\boldsymbol{\theta}_{-i}} \ell(\boldsymbol{\theta}_i, \boldsymbol{\theta}_{-i}), \tag{2.95}$$

which represents the profile of the log-likelihood surface viewed from the perspective of the axis of $\boldsymbol{\theta}_i$.

The properties of profile likelihood estimators are known and generally summarized by the results in Theorem 2.6; for more details on likelihood inference, see Cruz *et al.* (2015, chapter 7).

Theorem 2.6 *Given $x_1, x_2, \ldots, x_n$ independent realizations of random variables $X_i \sim F(x)$ for all $i \in \{1, 2, \ldots, n\}$, and let $\widehat{\boldsymbol{\theta}}$ denote the MLE estimator for the model parameters $\boldsymbol{\theta} \in \mathbb{R}^d$. Then the profile-likelihood-based deviance statistic for the sub-parameter vector $\boldsymbol{\theta}_i \subseteq \boldsymbol{\theta}$ is distributed according to a chi-squared distribution as follows:*

$$D_p(\boldsymbol{\theta}_i) = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell_p(\widehat{\boldsymbol{\theta}}_i)\} \sim \chi^2_s, \tag{2.96}$$

where the degrees of freedom corresponds to the dimension of the sub-parameter vector $\boldsymbol{\theta}_i \in \mathbb{R}^s$, for $s \leq d$.

This statistical result can be utilized in the GEV parameter estimation context along with the profile likelihood formulation to derive statistical confidence intervals for the return levels in Definition 2.24.

Definition 2.24 *The return levels of the GEV distribution are known analytically as the functions of the model parameters, for a given quantile level of the GEV distribution* $H(x_p; \mu, \sigma, \gamma) = \mathbb{P}\mathrm{r}[X \leq x_p] = 1 - p$*, according to*

$$x_p = \begin{cases} \mu - (\sigma/\gamma)\left[1 - \{-\ln(1-p)\}^{-\gamma}\right], & \gamma \neq 0, \\ \mu - \sigma \ln(-\ln(1-p)), & \gamma = 0, \end{cases} \tag{2.97}$$

where in common terminology x_p *is the return level associated with return period 1/p.*

The return level derives its name since to a reasonable degree of accuracy, the level x_p will typically be exceeded on average once every $1/p$ periods (years etc.). As detailed in Coles (2001), to obtain these confidence intervals for the return level under the profile likelihood, one reparameterizes the model according to the change of variable given by

$$\mu = x_p + \frac{\sigma}{\gamma}[1 - (-\ln(1-p))^{-\gamma}], \tag{2.98}$$

to obtain the likelihood in terms of the return level x_p and the parameters σ and γ, i.e. (x_p, σ, γ).

2.4.3 GEV SUB-FAMILY APPROACHES IN OpRisk LDA MODELING

Next we detail several GEV sub-family approaches in OpRisk LDA modeling. This can be considered as an alternative approach to utilizing the GEV family in OpRisk settings, in the sense that instead of performing estimation of the EVI and model parameters from the data, a risk manager starts from the assumption that the data is generated from a particular parametric member of the GEV family of severity models. That is, one considers different parametric models from this family, which admit densities and to work within the constraints of these sub-families, fitting them to the observed data via MLE, MOM or mixed-based approaches and then performing model selection on the most appropriate model. This could be achieved via likelihood ratio tests or penalized methods such as Akaike information criterion (AIC) or corrected Akaike information criterion (AICc), or in the Bayesian setting via Bayes factors, Bayesian information criterion or deviance information criterion's. Such topics are discussed in detail in Cruz *et al.* (2015, chapters 7 and 8). Next we provide a selection of detailed coverage of important parametric member sub-families of the GEV severity model for each EVT type.

2.4.4 PROPERTIES OF THE FRECHET–PARETO FAMILY OF SEVERITY MODELS

As the focus of this chapter is on heavy-tailed severity models in OpRisk, we provide the following results related to the expected loss in Proposition 2.5 for distributions in the Frechet domain $F \in \mathscr{D}(M)$ with EVI $\gamma > 0$.

Proposition 2.5 *A severity distribution* F *that satisfies Condition* D_I *in Definition 2.16 or the related conditions 2.3 in Proposition 2.3, which has a maximal limiting distribution* $F \in \mathscr{D}(M)$*, for which the EVI* $\gamma > 0$*, has finite pth moment as a function of the EVI magnitude according to*

$$\mathbb{E}\left[X_+^p\right] = \begin{cases} \infty, & p\gamma > 1, \\ < \infty, & p\gamma < 1. \end{cases} \tag{2.99}$$

Other members of the Frechet–Pareto class of extremal distributions include the families of the F-distribution, inverse gamma distribution, the log-gamma distribution and the folded T-distributions, the g-and-h family of distributions and the stable distributions, which are describe extensively in Cruz *et al.* (2015). In the next two Sections 2.4.5 and 2.4.6, we present two practically relevant families of OpRisk severity models that are prominent members of the Frechet–Pareto class. We illustrate their properties, how to fit and simulate from such models as well as the properties of these models with respect to the EVI and its consequences on capital estimation in an LDA OpRisk framework.

2.4.5 SINGLE RISK LDA POISSON-GENERALIZED PARETO FAMILY

If the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N_t}$, is chosen to be either a Pareto (Type I) distribution $X_i \sim Pareto(\alpha)$ or a generalized Pareto $X_i \sim GPD(\sigma, \gamma)$, then the limiting distribution of the maximum loss will be of a Frechet–Pareto type and following properties will apply. The distributions for the losses in each case are members of the Hall class of distributions and are, respectively, detailed later. For more details on properties of this family of parametric severity models, see Hosking & Wallis (1987).

2.4.5.1 Properties of Severity Model: Strict Pareto. If $X_i \sim Pareto(\alpha)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be $1/\alpha$ and survival and quantile functions are given by

$$\overline{F}(x;\alpha) = 1 - F(x) = \begin{cases} \left(\frac{x_{\mathrm{m}}}{x}\right)^{\alpha}, & \text{for } x \geq x_{\mathrm{m}}, \\ 1, & \text{for } x < x_{\mathrm{m}}, \end{cases}$$

$$f_X(x;\alpha) = \begin{cases} \alpha \, \frac{x_{\mathrm{m}}^{\alpha}}{x^{\alpha+1}}, & \text{for } x \geq x_{\mathrm{m}}, \\ 0, & \text{for } x < x_{\mathrm{m}}, \end{cases}$$

$$\lim_{x\to\infty} \overline{F}(x) = x^{-\alpha} l_F(x), \text{ with } l_F(x) = 1, x_m = 1,$$

$$Q(p) = (1-p)^{-1/\alpha}.$$

In addition, it is practically useful for statistical estimation purposes to know the moments and moment generating function of this distribution given by

$$\mathbb{E}[X] = \frac{\alpha x_{\mathrm{m}}}{\alpha - 1}, \text{ for } \alpha > 1, \ \mathbb{V}\mathrm{ar}[X] = \left(\frac{x_{\mathrm{m}}}{\alpha - 1}\right)^2 \frac{\alpha}{\alpha - 2}, \text{ for } \alpha > 2$$

$$\mathbb{E}[X^n] = \frac{\alpha x_{\mathrm{m}}^n}{\alpha - n}, \text{ for } \alpha > n,$$

$$M(t, \alpha, x_{\mathrm{m}}) = \mathbb{E}[e^{tX}] = \alpha(-x_{\mathrm{m}}t)^{\alpha}\Gamma(-\alpha, -x_{\mathrm{m}}t) \text{ and } M(0, \alpha, x_{\mathrm{m}}) = 1.$$

These can then be utilized by practitioners to fit such models to data, should they wish to perform a moment-matching-based estimation.

2.4.5.2 Properties of Severity Model: Generalized Pareto. If $X_i \sim GPD(\mu, \sigma, \gamma)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be γ and survival and quantile functions are given by

$$F(x;\gamma,\mu,\sigma) = \begin{cases} 1-(1+\gamma(x-\mu)/\sigma)^{-1/\gamma} & \text{for } \gamma \neq 0, \\ 1-\exp(-(x-\mu)/\sigma) & \text{for } \gamma = 0. \end{cases}$$

$$\overline{F}(x;\gamma,\mu,\sigma) = 1-F(x) = \left(1+\frac{\gamma(x-\mu)}{\sigma}\right)^{-1/\gamma}, \; x>0; \; \sigma,\gamma>0$$

$$f_X(x;\gamma,\mu,\sigma) = \frac{\sigma^{1/\gamma}}{(\sigma+\gamma(x-\mu))^{1+1/\gamma}};$$

$$\lim_{x\to\infty} \overline{F}(x) = x^{-\alpha} l_F(x), \text{ with } l_F(x) = \left(\frac{\sigma}{\gamma}\right)^{1/\gamma}\left(1+\frac{\sigma}{\gamma x}\right)^{-1/\gamma},$$

$$\lim_{x\to\infty} U(x) = x^{\gamma} l_U(x), \text{ with } l_U(x) = C(1+Dx^{-\beta}(1+o(1))).$$

for some constants $C>0$, $D\in\mathbb{R}$ and $\beta>0$. Also $x \geq \mu$ when $\gamma \geq 0$; or $\mu \leq x \leq \mu - \sigma/\gamma$ when $\gamma < 0$, where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ the scale parameter and $\gamma \in \mathbb{R}$ the shape parameter.

Remark 2.12 *In OpRisk modeling settings, it is typical that one would consider the situations in which the support of the GPD covered* $\mathbb{R}^+$*; hence, one would typically be interested in only EVI values for* $\gamma \geq \mu \geq 0$.

In the case of the GPD, the moments can be obtained, as detailed in Hosking & Wallis (1987) by considering, for $\mu = 0$, that $\mathbb{E}[(1+\gamma X/\sigma)^r] = 1/(1-r\gamma)$ if $1-r\gamma > 0$, where the rth moment of X exists if $\gamma < 1/r$. When they exist, the mean, variance, skewness and kurtosis are, respectively,

$$\begin{aligned} \mathbb{E}[X] &= \frac{\sigma}{1-\gamma}, \quad \mathbb{V}\text{ar}(X) = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)}, \\ \mathbb{S}\text{kewness}(X) &= \frac{2(1+\gamma)(l-2\gamma)^{1/2}}{(1-3\gamma)}, \\ \mathbb{K}\text{urtosis}(X) &= \frac{3(1-2\gamma)(3+\gamma-2\gamma^2)}{(1-3\gamma)(l-4\gamma)} - 3. \end{aligned} \tag{2.100}$$

In general, the properties of the GPD will depend on the shape parameter γ:

1. The case in which $\gamma = 0$ will produce the exponential distribution with the right tail unbounded.
2. For $\gamma > 0$, the GPD right tail is unbounded and the distribution is heavy tailed, so that some moments do not exist. In practice, it is known that estimation under real data sets for OpRisk settings can produce infinite mean models when $\gamma \geq 1$, see Moscadelli (2004) and Nešlehová *et al.* (2006).
3. For $\gamma < 0$, the right tail is bounded in $x \in [0, -\sigma/\gamma]$. This case is of relevance to OpRisk if there were, for example, a risk control mechanism or policy in a financial institution which would automatically cap the maximum allowable loss.

In addition, the GPD model has the special property that it can also be defined with respect to excesses. That is, given a random variable $X \sim F(x; \gamma, \mu, \sigma)$ for $x > 0$, then the distribution of the conditional excesses given by transformed random variable $X - u|X > u$ for a threshold u is also GPD with the same shape parameter γ and new scale parameters $\sigma + \gamma u$. A consequence of this property is that the mean excess when $\gamma < 1$ is given by

$$\mathbb{E}[X - u|X > u] = \frac{\sigma + \gamma u}{1 - \gamma}, \tag{2.101}$$

which will be used in future sections to also motivate EVT modeling approaches via points over threshold constructions.

2.4.5.3 Simulation of Generalized Pareto Random Variates via Transformation. Simulation of a loss process from a Pareto and generalized Pareto severity distribution can be achieved via a transformation of standard random variates as follows.

Algorithm 2.2 (Simulation from GPD)

1. *Generate a uniform random number U uniformly distributed on $(0, 1]$, i.e. $U \sim Uniform(0, 1)$.*
2. *Apply the transformation*

$$X = \mu + \frac{\sigma(U^{-\gamma} - 1)}{\gamma} \sim GPD(\mu, \sigma, \gamma). \tag{2.102}$$

2.4.5.4 OpRisk Severity, Annual Loss cdfs and VaR versus EVI. In Figure 2.8, we plot the cdf of the generalized Pareto severity distribution for a range of parameter settings $\gamma \in \{0.1, 0.5, 1, 1.5, 5\}$ and $\mu = 0, \sigma = 1$. Then in Figure 2.9, we plot the resulting LDA compound process for random variable $Z = \sum_{i=1}^{N} X_i$ with $N \sim Poisson(\lambda = 5)$ and i.i.d. losses $X_i \sim GPD(\mu = 0, \sigma = 1, \gamma)$ for all $i \in \{0, 1, \ldots, N\}$ for a range of parameter settings $\gamma \in \{0.1, 0.5, 1, 5\}$.

2.4.5.5 Simulation of LDA Poisson-Generalized Pareto Annual Loss Compound Process. Simulation of the annual loss process for T years, from a Poisson-generalized Pareto LDA model can be achieved via a transformation of standard random variates as follows.

Algorithm 2.3

1. *Generate vector of realized annual loss counts $\boldsymbol{N}_{1:T} = \{N_1, N_2, \ldots, N_T\}$ by drawing from a Poisson distribution with rate λ, $N_t \sim Poisson(\lambda)$. This is undertaken for each random variate realization $N_t = n_t$ for each year of simulation $t \in \{1, 2, \ldots, T\}$ via one of several possible algorithms, such as*
   (a) *Set $L = \exp(-\lambda)$, $s = 0$ and $p = 1$.*
   (b) *Generate $U \sim Uniform(0, 1)$ and set* $\mathrm{p} = \mathrm{pU}$.
   (c) *While $p > L$: set $s = s + 1$ then generate $U \sim Uniform(0, 1)$ and set $p = pU$.*
   (d) *Set $n_t = s$.*

2. *For each t, generate n_t i.i.d. realizations of the loss $X_i(t) \sim GPD(\mu, \sigma, \gamma)$ via transformation according to the following*
   (a) *Generate a uniform random number U uniformly distributed on $(0, 1]$.*
   (b) *Apply the transformation*

$$X = \mu + \frac{\sigma(U^{-\gamma} - 1)}{\gamma} \sim GPD(\mu, \sigma, \gamma). \tag{2.103}$$

3. *Set the realized annual loss in each year t to $Z(t) = \sum_{i=1}^{N_t} X_i(t)$.*

Finally, in Figure 2.10, we plot the VaR for the capital of a single risk LDA model constructed with a Poisson-generalized-Pareto structure. In particular, we focus on the effect on the VaR at a range of quantiles versus EVI, that is, the heaviness of the tails of the severity model.

Remark 2.13 (Time-Varying Tail Behaviour for GPD Models) *Recent interest in the application of generalized Pareto models in financial applications have considered extensions of such models to incorporate time-varying tail behaviour. The premise for such models in an OpRisk context is to consider a time series of losses X_t, where at some high threshold u, it is assumed that $X_t|X_t < u$ is modeled by a finite mixture distribution, while for $X_t|X_t > u$, a GPD is considered with time-varying parameters γ_t and σ_t, see examples in Embrechts et al. (1997) and Lopes* et al. *(2011) of these dynamic splice models.*

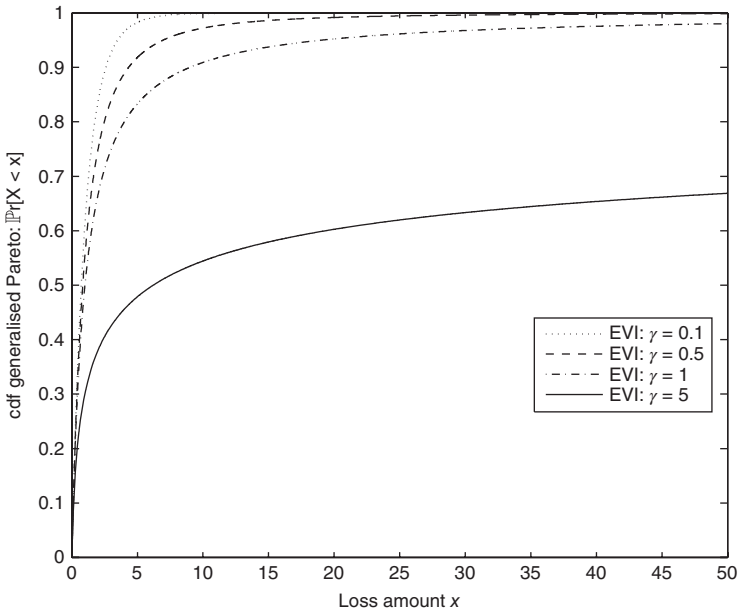

FIGURE 2.8 **Generalized Pareto severity distribution for different values of tail index EVI γ.**

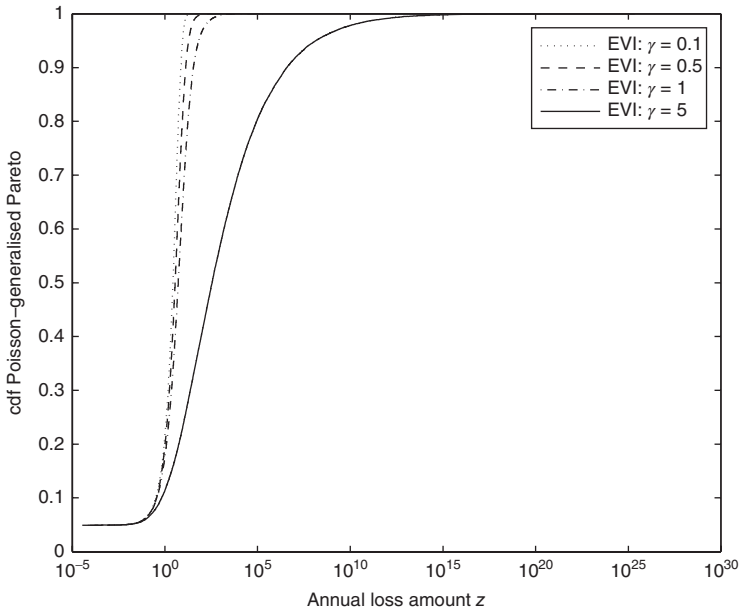

FIGURE 2.9 **Annual loss Poisson-generalized Pareto compound process LDA model for different values of tail index EVI γ.**

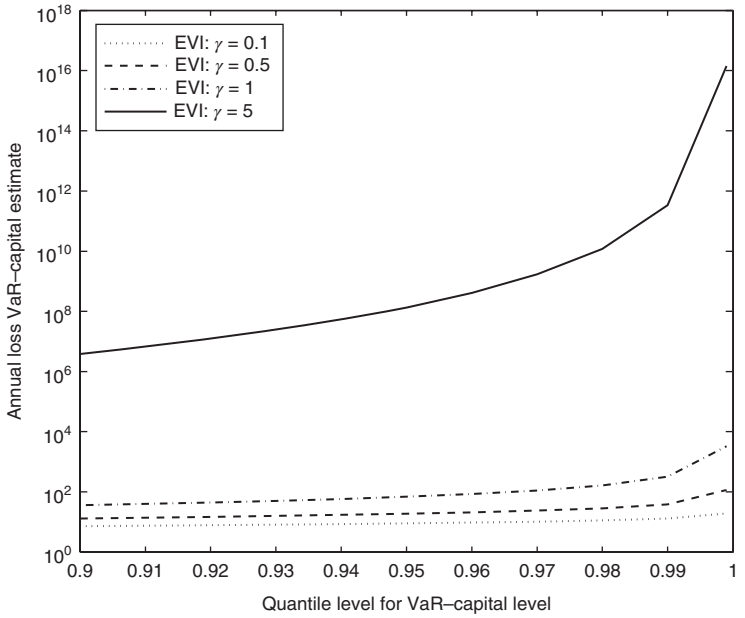

FIGURE 2.10 **Annual loss Poisson-generalized Pareto LDA model VaR for different values of tail index EVI γ.**

2.4.6 SINGLE RISK LDA POISSON-BURR FAMILY

In the next example, we consider some more flexible severity distribution examples proposed in the OpRisk literature in Dutta & Perry (2006), Peters & Sisson (2006a), which are also from the Hall class of distributions and which have as extremal limiting distributions for their maximum loss in the Frechet–Pareto class. In particular, we now consider examples of the most flexible members of the Burr family of distributions, see Burr (1942); Burr & Cislak (1968); Rodriguez (1977), which have been considered for the purpose of risk modeling and fitted as severity models in insurance and OpRisk models.

2.4.6.1 Properties of Severity Model: Burr Type XII. If the heavy-tailed severity model, selected to model the sequence of losses in year t, denoted by $\{X_i(t)\}_{i=1:n}$, is chosen to be a Burr Type XII distribution $X_i \sim BurrXII(\alpha)$, then the limiting distribution of the maximum loss will be of a Frechet–Pareto type and the following properties will apply.

The distribution for the losses in each case are members of the Hall Class of distributions and are given, respectively, in most basic form of the Burr Type XII family of distributions according to a distribution function

$$F_X(x;\tau,\lambda) = 1-(1+x^\tau)^{-\lambda}, \;\; x>0,\; \tau>0,\; \lambda>0, \tag{2.104}$$

where τ and λ are shape parameters. Introduction of location μ and scale η parameters can be achieved in numerous ways, such as for the scale case in one of the following ways:

$$\begin{aligned} F_X(x;\tau,\lambda,\eta) &= 1-\left(1+\frac{x^\tau}{\eta}\right)^{-\lambda}, \;\; x>0,\; \tau>0,\; \lambda>0,\; \eta>0, \\ F_X(x;\tau,\lambda,\eta) &= 1-\left(1+\left(\frac{x}{\eta}\right)^\tau\right)^{-\lambda}, \;\; x>0,\; \tau>0,\; \lambda>0,\; \eta>0. \end{aligned} \tag{2.105}$$

In the case in which $\eta = 1$, the density of the Burr XII severity model can be expressed according to

$$f_X(x;\tau,\lambda) = \lambda\tau x^{\tau-1}(1+x^\tau)^{-(\lambda+1)}, \;\; x\geq 0,\; \tau>0,\; \lambda>0. \tag{2.106}$$

Furthermore, if $X_i \sim BurrXII(\eta,\tau,\lambda)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be $(\lambda\tau)^{-1}$ and survival and quantile functions are given by

$$\begin{aligned} &\overline{F}(x;\eta,\tau,\lambda) = 1-F(x) = \left(\frac{\eta}{\eta+x^\tau}\right)^\lambda,\; x>0;\; \eta,\tau,\lambda>0 \\ &\lim_{x\to\infty}\overline{F}(x) = x^{-\lambda\tau}l_F(x),\; \text{with}\, l_F(x) = \left(\frac{\eta}{1+\eta/x^\tau}\right)^\lambda, \\ &\lim_{x\to\infty}U(x) = x^\gamma l_U(x),\; \text{with}\, l_U(x) = C(1+Dx^{-\beta}(1+o(1))), \end{aligned}$$

for slowly varying functions $l_F(x)$ and $l_U(x)$ and some constants $C>0$, $D\in\mathbb{R}$ and $\beta>0$.

In addition, it is practically useful for statistical estimation purposes to know the r-th moments of this family of distribution given by

$$\mathbb{E}[X^r] = \frac{\lambda\Gamma\left((\lambda-r)/\tau\right)\Gamma\left(1+r/\tau\right)}{\Gamma(\lambda+1)}. \tag{2.107}$$

These can then be utilized by practitioners to fit such models to data, should they wish to perform a moment-matching-based estimation. Note that restriction of $\lambda\tau > 4$ will ensure the existence of the fourth moment. In this case, it is interesting to consider the flexibility of this family of severity model for OpRisk by considering the possible skewness and kurtosis characteristics that are incorporated by the Burr Type XII family of models. In Tadikamalla (1980), these characteristics are presented diagrammatically for combinations of skewness and kurtosis (in Burr's notation (α_3, α_4)) and in the notation of Tadikamalla (1980) $(\mathbb{S}\text{kewness}[X], \mathbb{K}\text{urtosis}[X]) = (\sqrt{\beta_1}, \beta_2)$ given in Tadikamalla (1980, Figure 1).

Remark 2.14 *The regions of particular interest in the setting of OpRisk modeling are to consider restrictions to the parameters η, τ, λ of the Burr Type XII severity model which ensure that the skewness remains positive, that is, in the right quadrant of the range provided in Tadikamalla (1980, Figure 1).*

It is also worth noting the following members of the family of Burr Type XII severity models:

- Setting the shape parameter $\tau = 1$ recovers the Lomax distribution, see Lomax (1954).
- The exponential-gamma mixture of Dubey (1970) is equivalent to a Burr Type XII distribution in which $\beta \sim Gamma(\eta,\lambda)$ and the conditional severity distribution is given by $X|\beta \sim Exp(\beta)$ and then the marginal $X \sim ExpGamma(\eta, \lambda) = BurrXII(\tau = 1, \eta, \lambda)$.
- The Compound-Weibull mixture of Dubey (1970) can also be shown to be a form of Burr Type XII distribution, in which $\beta \sim Gamma(\tau, \lambda)$ and the conditional severity is given by $X|\beta \sim Weibull(\tau, \beta)$ and then the marginal $X \sim CompoundWeibull(\tau, \eta, \lambda) = BurrXII(\tau, \eta, \lambda)$ under a particular scaling representation.
- Finally, a practically important limiting cases of the Burr Type XII distribution that practitioners should consider is the lower bound on the skewness and kurtosis characteristics corresponding to the Weibull distribution, widely used in OpRisk LDA compound process models in practice.

2.4.6.2 Properties of Severity Model: Burr Type III. If the heavy-tailed severity model, selected to model the sequence of losses in year t, denoted by $\{X_i(t)\}_{i=1:n}$, is chosen to be a Burr Type III distribution $X_i \sim BurrIII(\alpha)$, then the limiting distribution of the maximum loss will also be of a Frechet–Pareto type and the following properties will apply.

The distribution for the losses in each case are members of the Hall Class of distributions and are given, respectively, in most basic form of the Burr Type III family of distributions according to a distribution function

$$F_X(x; \tau, \lambda) = (1 + x^{-\tau})^{-\lambda}, \quad x > 0, \ \tau > 0, \ \lambda > 0, \tag{2.108}$$

where τ and λ are shape parameters. Introduction of location μ and scale η parameters can be achieved in numerous ways as was the case for the Burr Type XII. The density of the Burr III severity model can be expressed according to the following equation:

$$f_X(x; \tau, \lambda) = \lambda\tau x^{-\tau-1}(1 + x^{-\tau})^{-\lambda-1}, \quad x \geq 0, \ \tau > 0, \ \lambda > 0. \tag{2.109}$$

If we then consider a scale parameter such that $X_i \sim BurrIII(\eta, \tau, \lambda)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be $1/\tau$ and survival and quantile functions are given by

$$\overline{F}(x;\eta,\tau,\lambda) = 1 - F(x) = 1 - \left(\frac{\eta}{\eta + x^{-\tau}}\right)^{\lambda},\ x > 0;\ \eta,\tau,\lambda > 0$$

$$\lim_{x\to\infty} \overline{F}(x) = x^{-\tau} l_F(x),\ \text{with}\, l_F(x) = \frac{\lambda}{\eta}\left(1 - \frac{\lambda+1}{2\eta}x^{-\tau} + o(x^{-\tau})\right),$$

$$\lim_{x\to\infty} U(x) = x^{\gamma} l_U(x),\ \text{with}\, l_U(x) = C(1 + Dx^{-\beta}(1 + o(1))),$$

for slowly varying functions $l_F(x)$ and $l_U(x)$ and some constants $C > 0$, $D \in \mathbb{R}$ and $\beta > 0$.

In addition, it is practically useful for statistical estimation purposes to know the rth moments of this family of distribution given for $\eta = 1$ by

$$\mathbb{E}[X^r] = \frac{\lambda\Gamma(\lambda + r/\tau)\,\Gamma(1 - r/\tau)}{\Gamma(\lambda + 1)}. \tag{2.110}$$

These can then be utilized by practitioners to fit such models to data, should they wish to perform a moment-matching-based estimation. Again we can present diagrammatically the flexibility of this family of severity model for OpRisk by considering the possible skewness and kurtosis characteristics that are incorporated by the Burr Type III family of models in Tadikamalla (1980, Figure 2).

Remark 2.15 *The regions of particular interest in the setting of OpRisk modeling are to consider restrictions to the parameters η, τ, λ of the Burr Type III severity model which ensure that the skewness remains positive, that is, in the right quadrant of the range provided in Tadikamalla (1980, Figure 2).*

2.4.6.3 Simulation of Burr Family Random Variates. Simulation of a loss process from a Burr Type XII and a Burr Type III severity distribution can be achieved straightforwardly via inversion, because as noted in Devroye & Devroye (1986, chapter 9, section 7.4) most members of the Burr family are unimodal. In addition, Pant & Headrick (2013) provide expressions in terms of the parameters that can be solved for the mode, when it exists, of the Type XII and Type III Burr distributions as functions of the quantile functions. The quantile functions are given by

$$\begin{aligned} Q_{XII}(u) &= \left((1-u)^{-1/\lambda} - 1\right)^{1/\tau},\ \lambda > 0,\ \tau > 0, \\ Q_{III}(u) &= \left(u^{-1/\lambda} - 1\right)^{1/\tau},\ \lambda > 0,\ \tau < 0. \end{aligned} \tag{2.111}$$

Then the mode of these distribution can be expressed in terms of the quantile functions as the solution to the non-linear equations, with $\widetilde{u}_{XII} = 1 - ((1+\tau\lambda)/(\tau+\tau\lambda))^{\lambda}$, $\widetilde{u}_{III} = ((1+\tau\lambda)/(\tau+\tau\lambda))^{\lambda}$, and

$$\begin{aligned} Q_{XII}(\widetilde{u}_{XII}) &= \left(\frac{\tau-1}{\tau\lambda+1}\right)^{1/\tau},\ \lambda > 0,\ \tau > 0, \\ Q_{III}(\widetilde{u}_{III}) &= \left(\frac{\tau-1}{\tau\lambda+1}\right)^{1/\tau},\ \lambda > 0,\ \tau < 0. \end{aligned} \tag{2.112}$$

Simulation of a loss process from a Burr Type III severity distribution can be achieved via a transformation of standard random variates as follows: see Pant & Headrick (2013, definition 2.2) where they provided the explicit functions for the quantiles of each family of models, given in Equation 2.111. The generation of a single random variate from each model is then achieved by the following steps.

Algorithm 2.4

1. *Generate a uniform random number U uniformly distributed on $(0, 1]$, i.e. $U \sim Uniform(0, 1)$.*
2. *Apply the transformation for BurrXII and BurrIII as follows*

$$X = Q_{XII}(U) = \left((1-U)^{-1/\lambda} - 1\right)^{1/\tau} \text{ and}$$
$$X = Q_{III}(U) = \left(U^{-1/\lambda} - 1\right)^{1/\tau}, \tag{2.113}$$

respectively.

2.4.6.4 Simulation of LDA Poisson-Burr Family Annual Loss Compound Process. Simulation of the annual loss process for T years, from a Poisson-Burr family (Type XII or Type III) LDA model can be achieved via a transformation of standard random variates as follows.

Algorithm 2.5

1. *Generate vector of realized annual loss counts $\boldsymbol{N}_{1:T} = \{N_1, N_2, \ldots, N_T\}$ by drawing from a Poisson distribution with rate λ, $N_t \sim Poisson(\lambda)$. This is undertaken for each random variate realization $N_t = n_t$ for each year of simulation $t \in \{1, 2, \ldots, T\}$ via one of several possible algorithms, such as the one used in Algorithm 2.3.*
2. *For each t, generate n_t i.i.d. realizations of the loss from either $X_i(t) \sim BurrXII(\tau, \lambda, \eta)$ or $X_i(t) \sim BurrIII(\tau, \lambda, \eta)$ via transformation according to the following:*
   1. *Generate a uniform random number U uniformly distributed on $(0, 1]$, i.e. $U \sim Uniform(0, 1)$.*
   2. *Apply the quantile function (inverse cdf) transformation:*

$$X = Q_{XII}(U) = \left((1-U)^{-1/\lambda} - 1\right)^{1/\tau}, \text{ or,}$$
$$X = Q_{III}(U) = \left(U^{-1/\lambda} - 1\right)^{1/\tau}. \tag{2.114}$$

3. *Set the realized annual loss in each year t to $Z(t) = \sum_{i=1}^{N_t} X_i(t)$.*

In Figure 2.11, the LDA model constructed with $N_t \sim Poisson(\lambda = 3)$ and severity distribution $X_i(t) \sim BurrXII(\tau, \lambda = 1, \eta = 1)$ is studied for values of EVI corresponding to $\gamma = (\lambda\tau)^{-1} \in \{0.1, 0.5, 1, 5\}$. In particular, we plot the annual loss cdf versus the EVI.

Finally, in Figure 2.12, we plot the VaR for the capital of a single risk LDA model constructed with a Poisson-generalized-Pareto structure. In particular, we focus on the effect on the VaR at a range of quantiles versus EVI, that is, the heaviness of the tails of the severity model.

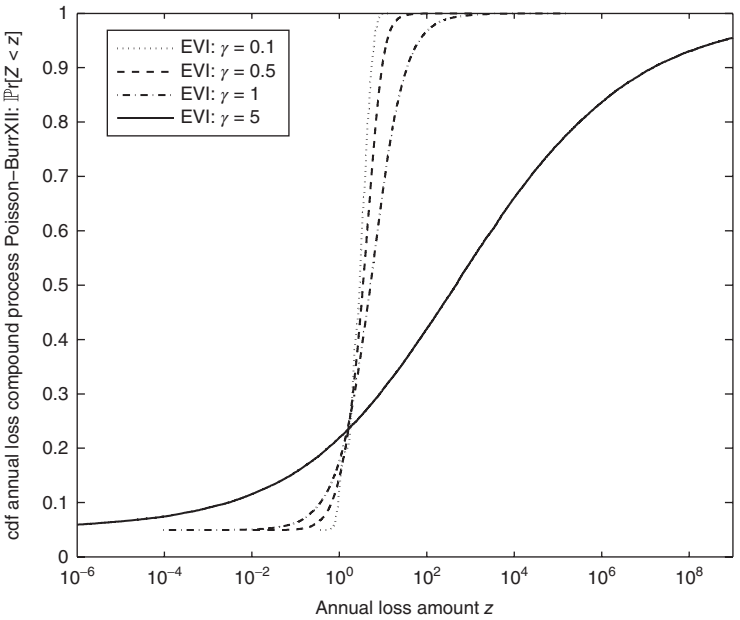

FIGURE 2.11 **Annual loss Poisson-Burr Type XII model for different values of tail index EVI $\gamma = (\lambda\tau)^{-1}$.**

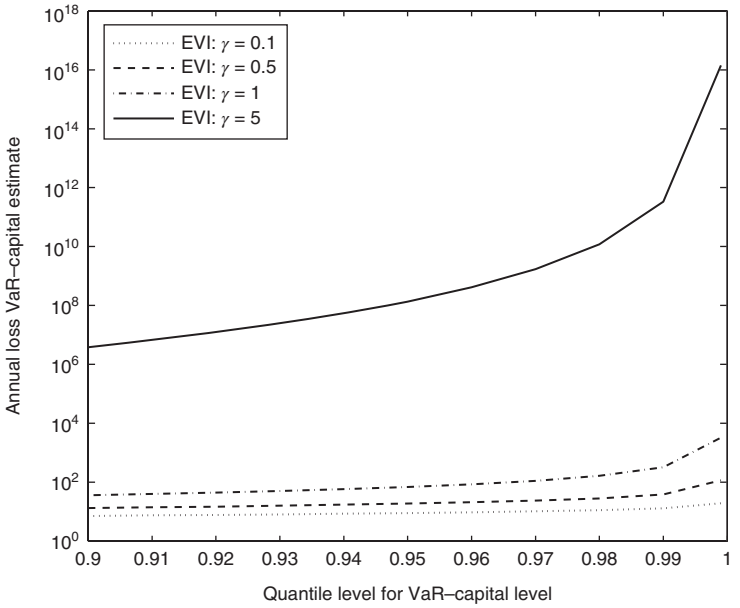

FIGURE 2.12 **Annual loss Poisson-generalized Pareto LDA model VaR for different values of tail index EVI γ.**

2.4.7 PROPERTIES OF THE GUMBEL FAMILY OF SEVERITY MODELS

As the focus of this chapter is on heavy-tailed severity models in OpRisk, we provide the following results related to the expected loss in Proposition 2.5 for distributions in the Gumbel domain $F \in \mathscr{D}(M)$ with EVI $\gamma = 0$.

The members of the Gumbel class of extremal distributions include the families of the Benktander II, Weibull, Exponential, Gamma, Logistic and LogNormal severity models. In the next section, we present two practically relevant families of OpRisk severity models that are prominent members of the Gumbel class. We illustrate their properties, how to fit and simulate from such models as well as the properties of these models with respect to the EVI and its consequences on capital estimation in an LDA OpRisk framework.

2.4.8 SINGLE RISK LDA POISSON-LogNormal FAMILY

If the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N}$, is chosen to be a LogNormal distribution $X_i \sim LogNormal(\mu, \sigma^2)$, then the limiting distribution of the maximum loss will be of a Gumbel type and following properties will apply. The two parameters in this model correspond to parameterizing the shape of the distribution for the severity σ and the log scale of the distribution μ.

2.4.8.1 Properties of Severity Model: LogNormal. If $X_i \sim LogNormal(\mu, \sigma^2)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be $\gamma = 0$ and survival and quantile functions are given by

$$\begin{aligned}
\overline{F}(x; \mu, \sigma) &= 1 - F(x) \\
&= \int_x^\infty \frac{1}{u\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\ln(u) - \mu)^2\right) du \\
&= \frac{1}{2} + \frac{1}{2}\ \mathrm{erf}\left[\frac{\ln(x) - \mu}{\sqrt{2\sigma^2}}\right],\ x > 0;\ \mu \in \mathbb{R}\ \sigma > 0
\end{aligned}$$

$$f_X(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}}\ \exp\left(-\frac{1}{2\sigma^2}(\ln(x) - \mu)^2\right),\ x > 0;\ \mu \in \mathbb{R}\ \sigma > 0$$

$$Q(p) = \exp(\mu + \sigma\Phi^{-1}(p)),\ 0 < p < 1.$$

In addition, it is practically useful for statistical estimation purposes to know the moments and moment-generating function of this distribution given by

$$\mathbb{E}[X] = e^{\mu+\sigma^2/2},\ \ \mathbb{V}\mathrm{ar}(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2},$$

$$\mathbb{E}[X^n] = \exp\left(n\mu + \frac{1}{2}n^2\sigma^2\right),\ \forall n \in \mathbb{J}^+.$$

These can then be utilized by practitioners to fit such models to data, should they wish to perform a moment-matching-based estimation. In addition, the following analytic expressions

for the summary statistics of this distribution, with respect to the parameters of the model, are known:

$$\mathbb{M}\text{edian}[X] = e^{\mu}, \;\; \text{Mode}[X] = e^{\mu - \sigma^2},$$

$$\mathbb{S}\text{kewness}[X] = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}.$$

2.4.8.2 Simulation of LogNormal Random Variates via Transformation and Box–Muller Approach. Simulation of a loss process from a LogNormal severity distribution can be achieved via a transformation of a Gaussian random variate, which itself is generated typically via a transformation of two uniform random variates in a Box–Muller approach, detailed as follows.

Algorithm 2.6

1. *Generate two independent uniform random numbers U_1, U_2 uniformly distributed on* $(0, 1]$
2. *Apply the Box–Muller transformation*

$$Y = \mu + \sigma \left(\sqrt{12 \ln(U_2)} \cos(2\pi U_1) \right) \sim \textit{Normal}(\mu, \sigma^2). \tag{2.115}$$

3. *Apply the transformation from Normal to LogNormal*

$$X = \exp(Y) \sim \textit{LogNormal}(\mu, \sigma^2). \tag{2.116}$$

2.4.8.3 OpRisk Severity, Annual Loss cdfs and VaR versus Shape. In Figure 2.13, we plot the cdf of the LogNormal severity distribution for a range of parameter settings $\sigma \in \{0.1, 0.5, 1, 5\}$ and $\mu = 1$. Then in Figure 2.14, we plot the resulting LDA compound process for random variable $Z = \sum_{i=1}^{N} X_i$ with $N \sim \textit{Poisson}(\lambda = 5)$ and i.i.d. losses $X_i \sim \textit{LogNormal}(\mu = 1, \sigma^2)$ for all $i \in \{0, 1, \ldots, N\}$ for a range of parameter settings $\sigma \in \{0.1, 0.5, 1, 5\}$.

2.4.8.4 Simulation of LDA Poisson-LogNormal Annual Loss Compound Process. Simulation of the annual loss process for T years from a Poisson-LogNormal LDA model can be achieved via a transformation of standard random variates as follows.

Algorithm 2.7

1. *Generate vector of realized annual loss counts $\boldsymbol{N}_{1:T} = \{N_1, N_2, \ldots, N_T\}$ by drawing from a Poisson distribution with rate λ, $N_t \sim \textit{Poisson}(\lambda)$. This is undertaken for each random variate realization $N_t = n_t$ for each year of simulation $t \in \{1, 2, \ldots, T\}$ via one of several possible algorithms, such as the one used in Algorithm 2.3*
2. *For each t, generate n_t i.i.d. realizations of the loss $X_i(t) \sim \textit{LogNormal}(\mu, \sigma^2)$ via transformation according to the following:*
   (a) *Generate two independent uniform random numbers U_1, U_2 uniformly distributed on* $(0, 1]$

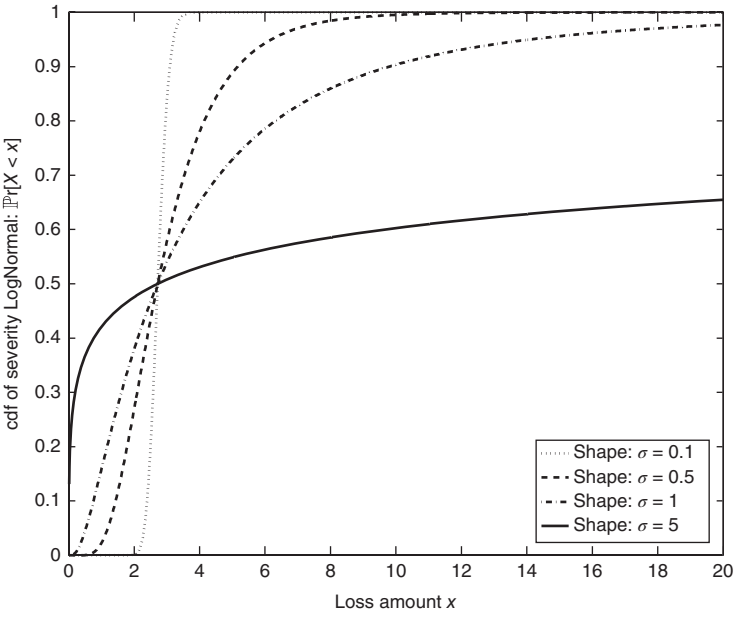

FIGURE 2.13 LogNormal severity distribution for different values of shape parameter σ.

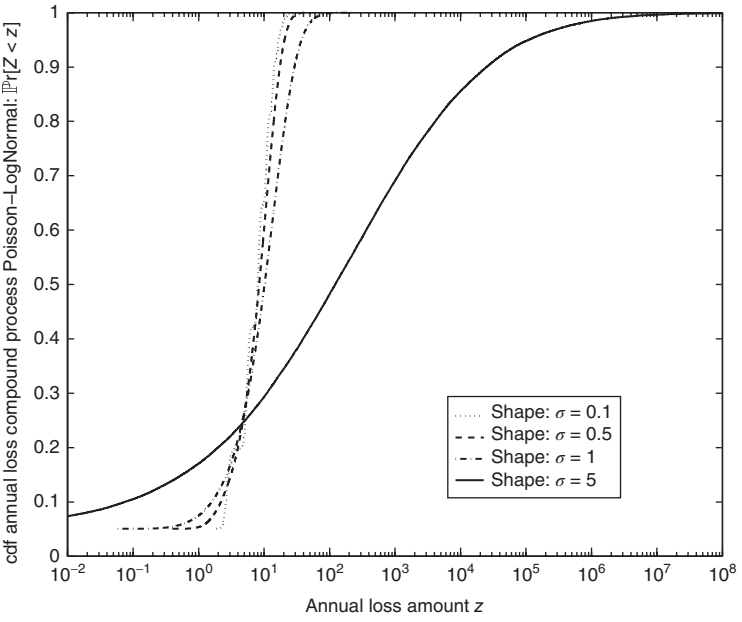

FIGURE 2.14 Annual loss Poisson-LogNormal compound process LDA model for different values of shape parameter σ.

(b) Apply the Box–Muller transformation

$$Y = \mu + \sigma\left(\sqrt{12\ln(U_2)}\cos(2\pi U_1)\right) \sim Normal\,(\mu, \sigma^2). \tag{2.117}$$

(c) Apply the transformation from Normal to LogNormal

$$X = \exp(Y) \sim LogNormal(\mu, \sigma^2). \tag{2.118}$$

3. *Set the realized annual loss in each year t to $Z(t) = \sum_{i=1}^{N_t} X_i(t)$.*

Finally, in Figure 2.15, we plot the VaR for the capital of a single risk LDA model constructed with a Poisson-LogNormal structure. In particular, we focus on the effect on the VaR at a range of quantiles versus shape parameter σ, that is, the heaviness of the tails of the severity model, for $\lambda = 5$ and $\mu = 1$.

2.4.9 SINGLE RISK LDA POISSON-BENKTANDER II MODELS

If the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N}$, is chosen to be a Benktander II distribution $X_i \sim BII(\mu, \sigma)$, then the limiting distribution of the maximum loss will be of a Gumbel type and following properties will apply. The two parameters in this model correspond to parameterizing the shape of the distribution for the severity α and the scale of the distribution β. This family of severity models actually has two variants, the Benktander I (Benktander–Gibrat) and Benktander II (Benktander–Weibull). The Benktander II family has enjoyed a wide variety of applications in

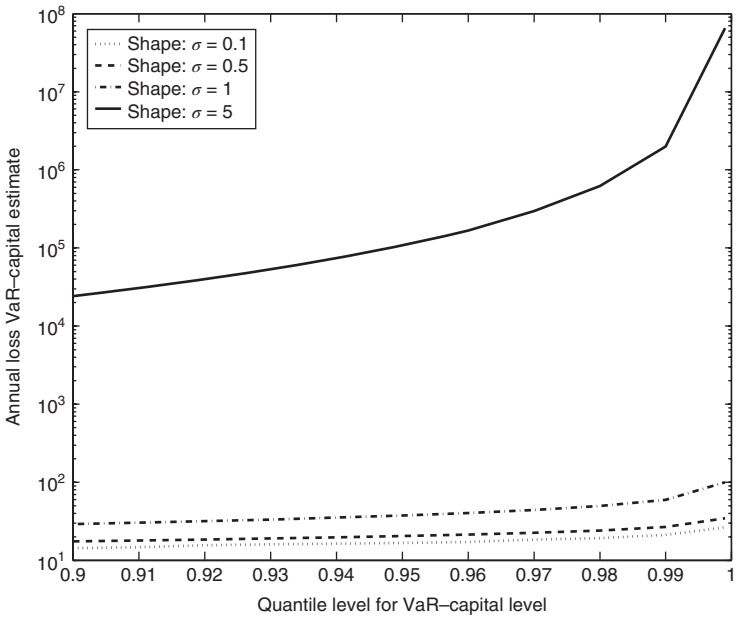

FIGURE 2.15 **Annual loss Poisson-LogNormal LDA model VaR for different values of shape σ.**

the literature of insurance (liability business and claims reserving applications) and risk modeling, because it was first proposed in the context of statistical parametric models for claims reserving applications in Benktander & Segerdahl (1960).

The asymptotic behaviour of the tail of the Benktander II distribution can be shown to lie between the exponential distribution and the Pareto distribution and is parameterized by two parameters which can be interpreted as shape α and scale β. In the limiting case in which the shape parameter approaches the lower limit of its domain from the right, that is, $\alpha \to 0$, then one recovers the Pareto distribution. In the case in which one sets the shape parameter $\alpha = 1$ and the scale parameter $\beta + 1 = \lambda$, then one recovers the exponential distribution.

2.4.9.1 Properties of Severity Model: Bektander II. If $X_i \sim BII(\alpha, \beta)$, then the extremal index of the limiting distribution for $X_{(n)}$ will be $\gamma = 0$ and survival and quantile functions, see 'Mathematica - Benktander Weibull Function', are given by

$$\overline{F}(x;\alpha,\beta) = 1 - F(x) = \exp\big(\beta(1-x^\alpha)/\alpha\big)x^{-1+\alpha},\ x > 1;\ \alpha \in (0,1]\ \beta > 0$$

$$f_X(x;\alpha,\beta) = \exp\big(\beta(1-x^\alpha)/\alpha\big)x^{-2+\alpha}(1-\alpha+\beta x^\alpha)$$

$$Q(p) = \begin{cases} 1 - \dfrac{\ln(1-p)}{\beta},\ 0 < p < 1,\ \alpha = 1, \\ \left(\left(\dfrac{1-\alpha}{\beta}\right)\mathscr{W}\left[\dfrac{\beta\exp\left(\frac{\beta}{1-\alpha}\right)(1-p)^{-\frac{\alpha}{1-\alpha}}}{1-\alpha}\right]\right)^{1/\alpha},\ 0 < p < 1,\ \alpha \in (0,1), \end{cases}$$

where $\mathscr{W}[x]$ is the 'Lambert W function' (also called *Omega function* or *product logarithm function*) and is a set of functions corresponding to the branches of the inverse relation of the function $f(w) = w\exp(w)$, giving the expression for $\mathscr{W}[x]$ as

$$x = W(x)e^{W(x)} \tag{2.119}$$

for any complex number x. In general, we can also utilize a second component $\mathscr{W}[x,k]$ where $k = 1$ returns the principle solution for w in $x = w\exp(w)$, evaluated numerically via, for example, the Mathematica function 'ProductLog(x)' and k different from one returns the k-th solution.

In addition, it is practically useful for statistical estimation purposes to know the moments of the random loss with this distribution, given in 'Mathematica - Benktander Weibull Function' by

$$\mathbb{E}[X] = 1 + \frac{1}{\beta},\ \ \mathbb{V}\mathrm{ar}(X) = \beta^{-2}\left(-1 + \frac{2\beta}{\alpha}\exp\left(\frac{\beta}{\alpha}\right)\mathscr{E}\left[1 - \frac{1}{\alpha}, \frac{\beta}{\alpha}\right]\right),$$

where $\mathscr{E}[n,x]$ denotes the generalized exponential integral, otherwise known as the *exponential integral function* (or 'Misra function'), given by

$$\mathscr{E}[n,x] = \int_1^\infty \frac{e^{-xt}}{t^n}\ dt = x^{n-1}\Gamma(1-n,x). \tag{2.120}$$

In addition, one can obtain an analytic functional expression for the nth order moments of the Benktander II severity distribution, given in 'Mathematica - Benktander Weibull Function' by

$$\mathbb{E}[X^n] = \exp\left(\frac{\beta}{\alpha}\right)\left(\frac{\beta}{\alpha}\mathscr{E}\left[-1+\frac{1-n}{\alpha},\ \frac{\beta}{\alpha}\right] - \left(1-\frac{1}{\alpha}\right)\mathscr{E}\left[\frac{1-n}{\alpha},\ \frac{\beta}{\alpha}\right]\right),$$
$$\forall n \in \mathbb{J}^+.$$

These can then be utilized by practitioners to fit such models to data, should they wish to perform a moment-matching-based estimation.

The following analytic expressions for the summary statistics of this distribution, with respect to the parameters of the model are known, which would allow one to also perform moment matching given sample estimates of the skew and kurtosis of the loss process.

$$\mathbb{M}\text{edian}[X] = \begin{cases} 1+\dfrac{\ln 2}{\beta}, & \text{if } \alpha = 1, \\ \left(\dfrac{1-\alpha}{\beta}\mathscr{W}\left[\dfrac{2^{\frac{\alpha}{1-\alpha}}\beta\exp(\beta/(1-\alpha))}{1-\alpha}\right]\right)^{1/\alpha}, & \text{otherwise,} \end{cases}$$

$$\begin{aligned} \mathbb{S}\text{kewness}[X] = &-\frac{2}{\alpha}\left\{6\beta\exp\left(\frac{\beta}{\alpha}\right)\left(\beta\mathscr{E}\left[1-\frac{2}{\alpha},\ \frac{\beta}{\alpha}\right]\right)\right. \\ &\left.-(1+\beta)\mathscr{E}\left[1-\frac{1}{\alpha},\ \frac{\beta}{\alpha}\right]\right\} \\ &\times\left\{-1+\frac{2\beta}{\alpha}\exp\left(\frac{\beta}{\alpha}\right)\mathscr{E}\left[1-\frac{1}{\alpha},\ \frac{\beta}{\alpha}\right]\right\}^{-3/2}, \end{aligned}$$

$$\begin{aligned} \mathbb{K}\text{urtosis}[X] = &\left\{-3+\frac{12\beta}{\alpha}\exp\left(\frac{\beta}{\alpha}\right)\left[\beta^2\mathscr{E}\left[1-\frac{3}{\alpha},\ \frac{\beta}{\alpha}\right]\right.\right. \\ &\left.\left.+(1+\beta)\left(-2\beta\mathscr{E}\left[1-\frac{2}{\alpha},\ \frac{\beta}{\alpha}\right]+(1+\beta)\mathscr{E}\left[1-\frac{1}{\alpha},\ \frac{\beta}{\alpha}\right]\right)\right]\right\} \\ &\times\left\{1-\frac{2\beta}{\alpha}\exp\left(\frac{\beta}{\alpha}\right)\mathscr{E}\left[1-\frac{1}{\alpha},\ \frac{\beta}{\alpha}\right]\right\}^{-2}. \end{aligned}$$

2.4.9.2 Simulation of Benktander II Random Variates via Inversion. Simulation of a loss process from a Benktander II severity distribution can be achieved via a transformation of a uniform random variate and utilization of the knowledge of the quantile function, detailed as follows.

Algorithm 2.8

1. *Generate an independent uniform random number U uniformly distributed on* $(0,1]$
2. *Apply the quantile function (inversion technique) to obtain a draw from the Benktander II distribution*

$$X = \begin{cases} 1-\dfrac{\ln(1-U)}{\beta}, & \alpha = 1, \\ \left(\left(\dfrac{1-\alpha}{\beta}\right)\mathscr{W}\left[\dfrac{\beta\exp(\beta/(1-\alpha))\,(1-U)^{-\alpha/(1-\alpha)}}{1-\alpha}\right]\right)^{1/\alpha}, & \alpha \in (0,1), \end{cases} \tag{2.121}$$

where $X \sim BII(\alpha,\beta)$.

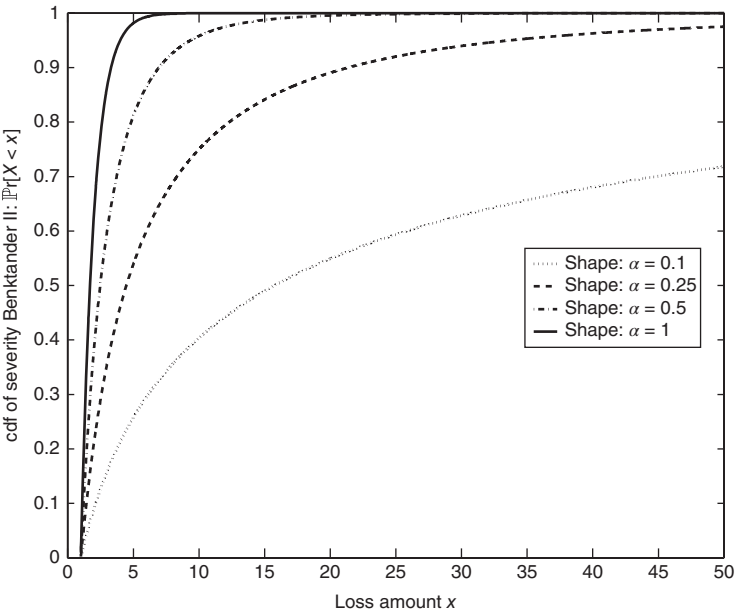

FIGURE 2.16 **Benktander II severity distribution for different values of shape parameter α.**

2.4.9.3 OpRisk Severity, Annual Loss cdfs and VaR versus Shape. In Figure 2.16, we plot the cdf of the Benktander II severity distribution for a range of parameter settings $\alpha \in \{0.1, 0.25, 0.5, 1\}$ and $\sigma = 1$. Then in Figure 2.17, we plot the resulting LDA compound process for random variable $Z = \sum_{i=1}^{N} X_i$ with $N \sim Poisson(\lambda = 5)$ and i.i.d. losses $X_i \sim BII(\alpha, \beta = 1)$ for all $i \in \{0, 1, \ldots, N\}$ for a range of parameter settings $\sigma \in \{0.1, 0.25, 0.5, 1\}$.

2.4.9.4 Simulation of LDA Poisson-Benktander II Annual Loss Compound Process. Simulation of the annual loss process for T years, from a Poisson-LogNormal LDA model can be achieved via a transformation of standard random variates as follows.

Algorithm 2.9

1. *Generate vector of realized annual loss counts* $\boldsymbol{N}_{1:T} = \{N_1, N_2, \ldots, N_T\}$ *by drawing from a Poisson distribution with rate* λ, $N_t \sim Poisson(\lambda)$. *This is undertaken for each random variate realization* $N_t = n_t$ *for each year of simulation* $t \in \{1, 2, \ldots, T\}$ *via one of several possible algorithms, such as in Algorithm 2.3.*
2. *For each* t, *generate* n_t *i.i.d. realizations of the loss* $X_i(t) \sim BII(\alpha, \beta)$ *via inversion according to the following:*
   (a) *Generate an independent uniform random number* U *uniformly distributed on* $(0, 1]$.
   (b) *Apply the quantile function (inversion technique) to obtain a draw from the Benktander II distribution*

$$X_i(t) = \begin{cases} 1 - \dfrac{\ln(1-U)}{\beta}, & \alpha = 1, \\ \left(\left(\dfrac{1-\alpha}{\beta} \right) \mathscr{W} \left[\dfrac{\beta \exp\left(\frac{\beta}{1-\alpha} \right) (1-U)^{-\frac{\alpha}{1-\alpha}}}{1-\alpha} \right] \right)^{\frac{1}{\alpha}}, & \alpha \in (0,1), \end{cases} \tag{2.122}$$

where $X_i(t) \sim BII(\alpha, \beta)$.

3. *Set the realized annual loss in each year* t *to* $Z(t) = \sum_{i=1}^{N_t} X_i(t)$.

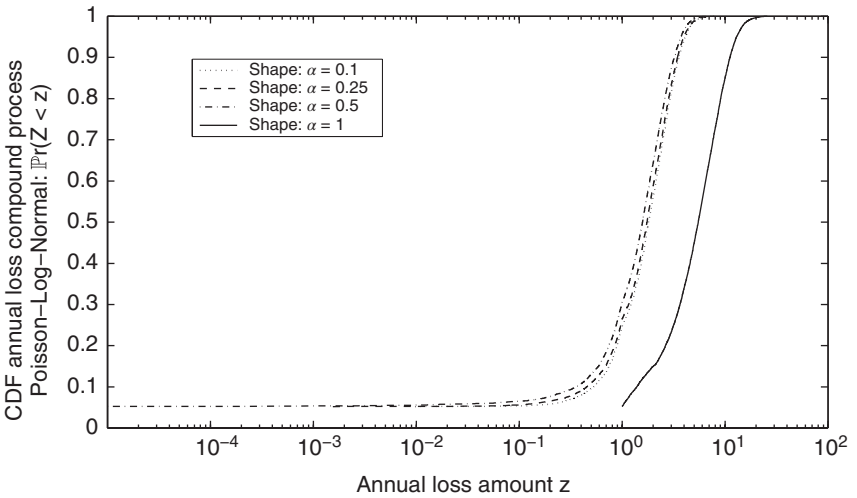

FIGURE 2.17 Annual loss Poisson-Benktander II compound process LDA model for different values of shape parameter α.

2.5 Theoretical Properties of Univariate EVT – Threshold Exceedances

In practical applications, one considers the EVT modeling approach as a tool which is used in one of two ways: to obtain a statistically accurate and robust best possible estimate of the tail area of the severity distribution or in an aggregate setting the actual loss distribution; or, as will be the case with many rare but high consequence loss process scenarios in which there will be a distinct lack of useful historical data, EVT may be used to inform or guide appropriate model selection for the distribution of such extreme losses in a conservative approach to modeling.

Here we detail an alternative approach to understanding the distribution obtained for the extreme of a sequence of losses in a single risk process. This approach is based on threshold exceedance rather than block maxima and tends to be considered practically more useful in scenarios in which the number of observed losses upon which the modeling is based is

relatively few. To understand this perspective, consider the previous approach based on block maxima. Under such an approach to EVT, the risk modeller can decide upon a suitable block period on which to study the largest loss, for example, the largest monthly, quarterly or annual loss. The scale or resolution of the modeling will clearly depend on the total number of losses observed in the given time scale and the suitability of the assumptions of equal numbers of losses per block. This can be problematic for many practical OpRisk modeling scenarios and the approach described next formulates an alternative perspective to deriving the distribution of the maximum loss via an exceedances argument. Such an approach is known as the Peaks Over Threshold (POT) method. The resulting statistical model derived from the POT approach to EVT is that of the GPD model. We have introduced the reader to this model in the LDA framework as an important sub-family of the Frechet limit in the GEV family. In this section, we discuss the one-to-one correspondence between this model under a POT approach and the GEV model under the block maxima approach, this is formulated in an elegant theory known as the *Pickands–Balkema–deHaan theorem*. This result involves the study of the limiting distribution of points exceeding a threshold. In addition, we will detail fundamental results surveying the distribution of the number of exceedances above the threshold.

To proceed with an understanding of this approach, we first define the notion of exceedances, exceedance times and then finally the POT concept. An exceedance for a given set of observations $\{x_1, x_2, \ldots, x_n\}$ and a threshold u is given by Definition 2.25.

Definition 2.25 (Exceedances) *The jth observation in a set of i.i.d. observations $\{x_1, x_2, \ldots, x_n\}$, which correspond to realizations of i.i.d. losses $X_i \sim F(x)$ for all $i \in \{1, 2, \ldots, n\}$, is an exceedance with respect to a threshold level u if $x_j > u$ and the excess is defined as the difference $x_j - u$.*

It is then a natural question to try to formulate the distribution for the excess given by

$$F_{X>u}(x) = \mathbb{P}\mathrm{r}[X - u \leq x | X > u], \; 0 \leq x < x_* - u, \tag{2.123}$$

where we define $x_* = \sup\{x : F(x) < 1\}$ as the right end point of the severity distribution F. In particular, we may be interested in the extreme value modeling context on the existence and uniqueness of the limiting distribution as the threshold level u is asymptotically growing large, i.e. $u \to \infty$. This leads one to the fundamental result obtained when studying the limiting properties and distributional convergence of $F_{X>u}(x)$ for a general class of severity models that may have generated the loss process observed, given in Theorem 2.7.

Theorem 2.7 (Pickands–Balkema–deHaan Theorem) *Given losses from a risk process, random variables $\{X_1, X_2, \ldots, X_n\}$ which are assumed i.i.d. with severity distribution F, the following are equivalent:*

1. *The severity distribution F is a member of the maximal domain of attraction of the GEV distribution $F \in \mathcal{D}(H_\gamma)$.*

2. *There exists a measurable function $\beta(\cdot)$ such that the following limit is satisfied, for the Generalized Pareto Distribution, see Section 2.4.5,*

$$\lim_{u \to x_*} \sup_{0 \leq x < x_* - u} \left| F_{X>u}(x) - GPD\left(\frac{x}{\beta(u)}; \gamma\right) \right| = 0. \tag{2.124}$$

This result shows that one may now start to consider a statistical model for the number of observed losses from the risk process that exceed a high threshold u. In practical settings, this shifts the focus from simply estimating the tail quantile of a risk process to considering the distribution of the losses that exceed high thresholds such as a VaR quantile at a given level of significance.

2.5.1 UNDERSTANDING THE DISTRIBUTION OF THRESHOLD EXCEEDANCES

We will now go through a simple constructive argument to understand the result obtained for threshold exceedances in the univariate i.i.d. setting. This will be valuable as when more advanced topics relating to dependence are considered, we will utilize these fundamentals to derive limiting distributions in such cases. In considering the understanding of threshold exceedances and the POT's framework, it will be beneficial to recall some elementary properties of point processes.

A general point process can be defined informally as the set of points $t_i \in \mathcal{S}$ randomly distributed in a state space $\mathcal{S}$ which are indexed by a set $\mathcal{I}$, denoted by $\{t_i : i \in \mathcal{I}\}$. Given this process, a point process defined for this state space and point set corresponds to counting the number of points in regions of $\mathcal{S}$ which will produce for each subset $A \subseteq \mathcal{S}$ a random process defined by

$$N(A) = \sum_{i \in \mathcal{I}} \mathbb{I}(t_i \in A), \ A \subseteq \mathcal{S}. \tag{2.125}$$

Having defined the counting process on subsets of the state space, one can also define the expected number of counts over such subsets, known generically as the *intensity measure*, $\Lambda(A) = \mathbb{E}[N(A)]$. This is defined in topological spaces and more abstract settings, but it is sufficient to consider in this text cases in which the state space is Euclidean space or a subset of Euclidean space in dimension d, here with $d = 1$. In such settings, if the intensity measure admits a density, in other words the Radon–Nikodym derivative of the intensity measure exists, say with respect to Lebesgue measure on the state space, then we can state that

$$\Lambda(A) = \int_A \lambda(x)\, dx, \tag{2.126}$$

where $\lambda(x)$ is the 'intensity function' of the process. Such general specifications extend to many settings including multivariate spatial, temporal or spatial-temporal state spaces and homogeneous (constant intensity functions) and inhomogeneous (changing intensity functions) over the subsets $A \subseteq \mathcal{S}$ of the state space, see advanced discussions in Cox & Isham (1980), Kingman (1993), Jacobsen (2006) and Resnick (1986).

Note, in the context in which we consider such processes in this chapter, typically, the point process involves occurrences of losses at particular random times in a time interval $[0, T]$. It is particularly of interest to risk modeling to then consider not just the number of occurrences that have occurred in such an interval but also the amounts of such losses and, in particular, here the

amounts once they exceed some threshold level. This naturally leads one to the consideration of a special class of point processes, which is 'marked' at particular points in the state space either deterministically or randomly. Such a marked point process is defined according to the random variables

$$N(A) = \sum_{i \in \mathcal{F}} X_i \mathbb{I}(t_i \in A), \ A \subseteq \mathcal{S}. \tag{2.127}$$

Here, the marked point process counts for each t_i a quantity X_i (loss amount) in a given subset A, which could, for example, be an interval of time $[t_1, t_2] \in \mathcal{S} = [0, T]$, in which the marks X_i are identically distributed. Having defined this generic representation of a marked point process, we note that a compound Poisson process, such as those used throughout this book to model single loss process under an LDA framework, is simply a marked point process for which the points t_i occur according to a Poisson process independently of the marks (losses) X_i, which are themselves in the simplest settings assumed i.i.d. from a severity distribution $F(x)$.

In Cox & Isham (1980, section 1.3), the basic discussion on specifications of point processes is provided. In particular, if one considers only processes in which no multiple simultaneous occurrences occur in a given small subset A. That is, if one restricts to processes which satisfy the condition that for a small subset $A = [t, t + \delta]$, such as an infinitesimal interval of time in which $\delta \approx 0$, the following holds

$$\mathbb{P}\mathrm{r}[N(t, t + \delta) > 1] = o(\delta), \tag{2.128}$$

then, in general, the specification of a simple point process can be achieved by consideration of one of the following equivalent representations:

1. specification of the intensity function; or
2. specification of joint distribution of the intervals between successive points, via an interval specification; or
3. the finite dimensional distribution specification involving the joint distribution of the numbers of points in arbitrary sets $A_1, A_2, \ldots$, known as the *counting specification*.

Throughout this section, it will be useful to recall the definition of the special case of a point process or counting process and a renewal process defined over time, given in Definitions 2.26 and 2.27.

Definition 2.26 (Counting Process in Time) *A stochastic process* $\{N(t), t \geq 0\}$ *is a counting process if* $N(t)$ *represents the total number of 'events' that have occurred up to time* t*. Furthermore, define* $\{\tau_i, \ i \in \{1, \ldots, N(t)\}\}$ *to denote the set of times at which the events occurred and* $\{\widetilde{\tau}_i, \ i \in \{1, \ldots, N(t)\}\}$ *denotes the set of inter-arrival times,* $\widetilde{\tau}_1 = \tau_1 - 0$ *and* $\widetilde{\tau}_i = \tau_i - \tau_{i-1}$ *for* $i > 1$*, between event occurrences.*

Given the definition of a counting process, we can go onto define a renewal process under the interval specification of a point process.

Definition 2.27 (Renewal Process) *A renewal process is a counting process* $\{N(t), t \geq 0\}$ *for which the inter-arrival times* $\{\widetilde{\tau}_i, \ i \in \{1, \ldots, N(t)\}\}$ *are each i.i.d.*

If one makes a special choice of the distribution of inter-arrival times in the renewal process corresponding to exponentially distributed inter-arrivals, then the memoryless property holds

for waiting times. This is a special example of a renewal process corresponding to the simple temporal homogeneous Poisson process described in Example 2.3.

EXAMPLE 2.3

A counting process $\{N(t), t \geq 0\}$ is a homogeneous temporal Poisson process with rate λ, if the inter-arrival times $\widetilde{\tau}_i$ in the set $\{\widetilde{\tau}_i,\ i \in \{1, \ldots, N(t)\}\}$ are all i.i.d. with exponential distribution given by

$$\mathbb{P}\mathrm{r}[\widetilde{\tau}_1 \leq t] = 1 - \exp(-\lambda t),\ t \geq 0. \tag{2.129}$$

Furthermore, the distribution of the number of events in an interval of time $[0, t]$ is Poisson and given by

$$\mathbb{P}\mathrm{r}[N(t) = n] = \exp(-\lambda t)\frac{(\lambda t)^n}{n!},\ n \in \{0, 1, 2, \ldots\}. \tag{2.130}$$

Finally, the total waiting time until the nth event defined by random variable

$$S_n = \sum_{k=1}^{n} \widetilde{\tau}_k$$

is distributed according to the Erlang distribution given by a density

$$f_{S_n}(t) = \lambda \exp(-\lambda t)\frac{(\lambda t)^{n-1}}{(n-1)!},\ t \geq 0 \tag{2.131}$$

with an expected number of events in $[0, t]$ given by $\mathbb{E}[N(t)] = \lambda t$. ■

Bearing in mind the definitions of a renewal process, we return back to considerations of exceedances of the single risk processes losses above some threshold level. We consider a set of n losses and times at which these losses occurred grouped in tuples for a given interval of time $[0, T]$ denoted by $\{(X_i, t_i)\}_{i=1:n}$. We then define a secondary stochastic process derived from the first process $\{(X_i, t_i)\}_{i=1:n}$ over the index set of the losses from $\{0, 1, \ldots, n\}$, which will generate a subset of 'times of exceedances' of losses above a threshold, given by indexes which then naturally have an associated time of occurrence in the original process; hence, they are denoted as exceedance times. We begin with the definition of the exceedance times at which threshold exceedances occurred for a set of i.i.d. data $\{X_n\}_{n\geq 1}$ and a threshold level u, according to Definition 2.28.

Definition 2.28 *Given i.i.d. loss data $\{X_n\}_{n\geq 1}$ and a threshold level u, the exceedance times for the single risk process defined by these losses are random variables denoted by $\{\tau_j\}_{j\geq 1}$ and given by*

$$\tau_1 = \inf\{j \geq 1 :\ X_j > u\}$$

$$\tau_2 = \inf\{j \geq \tau_1 :\ X_j > u\}$$

$$\vdots\ \vdots$$

$$\tau_k = \inf\{j \geq \tau_{k-1} : \ X_j > u\}, \tag{2.132}$$

and the resulting exceedances indexed by these times correspond to the set of losses $\{X_{\tau_k}\}_{k\geq 1}$.

To illustrate the primary single risk process and the resulting derived secondary process developed above, consider the illustrative example provided in Figure 2.18.

EXAMPLE 2.4

In this illustrative example, we consider a single risk process in which the losses are recorded on an aggregated daily time scale over a year to specify a realization of the stochastic process $\{(X_i, t_i)\}_{i=1:n}$. It is assumed losses and loss times are i.i.d. and $X_i \sim LogNormal(\mu, \sigma^2)$ and $t_i \sim Exp(\mu)$, and the realization for a given year is depicted in Figure 2.18 by the open circles for losses not exceeding a threshold u and closed red circles for losses exceeding the threshold u, given there were $n = 50$ losses in the year (typically $N \sim Poisson(\lambda)$).

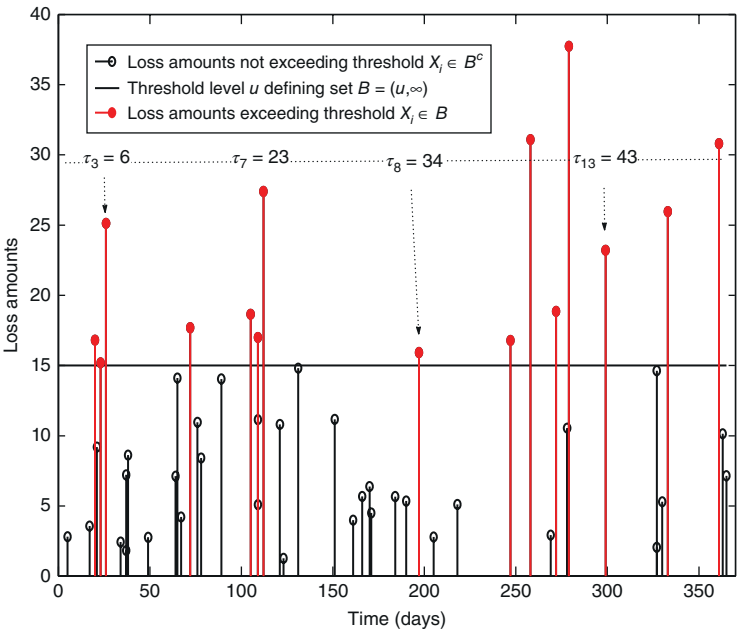

FIGURE 2.18 **An example realization of a single risk process $\{(X_i, t_i)\}_{i=1:n}$ with marked losses that exceeded a threshold loss amount u. The times on the x-axis at which the losses occurred correspond to the days t_i, the amounts correspond to the values X_i and the secondary process comprised of a subset of the losses, denoted by τ_j's as illustrated on the figure, corresponds to the indexes of the j-th exceedance.** ***See insert for color representation of this figure.***

Now following the constructive approach of (Resnick 2007, chapter 4), we can consider defining a random variable for the total number of exceedances of a threshold u in a set of i.i.d. loss

data $\{X_1, X_2, \ldots, X_n\}$ according to the random variable T_n given by

$$T_n = \sum_{j=1}^{n} \mathbb{I}_{(u,\infty)}(X_j), \tag{2.133}$$

which we can show has the distributional result given in Lemma 2.3.

Lemma 2.3 *Given i.i.d. loss data, $\{X_1, X_2, \ldots, X_n\}$, from a single risk process with severity distribution $F(x)$ and a threshold level u, the distribution of the random variable*

$$T_n = \sum_{j=1}^{n} \mathbb{I}_{(u,\infty)}(X_j), \tag{2.134}$$

is given by a binomial distribution according to

$$\Pr[T_n = k] = C_k^n (1 - F(u))^k F(u)^{(n-k)}, \ \forall k \in \{0, 1, \ldots, n\}. \tag{2.135}$$

To see this result, in the case in which we consider i.i.d. data from a fixed severity distribution $F(x)$ which is not time varying, then the distribution of exceedance is defined as the number of 'successes' in n Bernoulli trials, where a success is an exceedance and the probability of success is $p = \Pr[X > u] = 1 - F(u)$. Therefore, the distribution of the number of exceedances is simply binomial, $T_n \sim Binomial(n, p)$.

In addition, under the assumption of i.i.d. data for a given threshold level u, the exceedance times are i.i.d. random variables and the inter-arrival times between exceedances given by the difference times $\{\tau_1 - 0, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots\}$ are also i.i.d. random variables, which lead us to the distributional result in Lemma 2.4.

Lemma 2.4 *Consider i.i.d. loss data, $\{X_1, X_2, \ldots\}$, from a single risk process with severity distribution $F(x)$, a threshold level u and i.i.d. inter-arrival times between exceedances given by the difference times $\{\tau_1 - 0, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots\}$. Then the distribution of the random variable for the first exceedance time, τ_1, is given by a geometric distribution, $\tau_1 \sim Geometric(p)$ with $p = 1 - F(u)$, according to*

$$\Pr[\tau_1 = k] = (F(u))^{(k-1)}(1 - F(u)), \ \forall k \in \mathbb{J}^+, \tag{2.136}$$

where $\tau_1 = k$ refers to the kth loss being the first exceedance. Furthermore, the conditional distribution of the jth exceedance time, given the previous $j-1$ exceedance times, $\tau_j | \tau_1, \ldots, \tau_{j-1}$ will also be distributed according to a geometric distribution in the i.i.d. case defined by

$$\Pr[\tau_j = \widetilde{k} | \tau_{j-1} = m] = (F(u))^{(\widetilde{k}-1-m)}(1 - F(u)), \ \forall \widetilde{k} \in \{m+1, m+2, \ldots\}, \tag{2.137}$$

where $\tau_j = \widetilde{k} = k + m$ refers to the k-th loss after the $(j-1)$-th exceedance which occurred at the mth loss.

To see how this is obtained, consider the first exceedance time, it is well defined under these assumptions according to the probabilities

$$\begin{aligned} \Pr[\tau_1 > k] &= \Pr[X_1 \leq u, X_2 \leq u, \ldots, X_k \leq u] = F(u)^k, \\ \Pr[\tau_1 = k] &= \Pr[X_1 \leq u, X_2 \leq u, \ldots, X_k > u] = (F(u))^{(k-1)}(1 - F(u)). \end{aligned} \tag{2.138}$$

Therefore, we see from this analysis that the distribution of the first exceedance time is a geometric distribution with $\tau_1 \sim Geometric(p)$ where $p = 1 - F(u)$. In addition, as a result of the memoryless property of the geometric distribution, meaning that one can truncate and renormalize the distribution at any positive integer value and retain a geometric distribution, we can also state the conditional distributions of exceedance times using this property. Therefore, it is also true that the conditional distribution of the jth exceedance time, given the previous $j-1$ exceedance times, $\tau_j|\tau_1, \ldots, \tau_{j-1}$ defined by

$$\begin{aligned}
\Pr[\tau_j > k+m|\tau_{j-1} = m] &= \Pr[X_{\tau_{j-1}+1} \leq u, X_{\tau_{j-1}+2} \leq u, \ldots, X_{\tau_j} \leq u] = F(u)^k, \\
\Pr[\tau_j = k+m|\tau_{j-1} = m] &= \Pr[X_{\tau_{j-1}+1} \leq u, X_{\tau_{j-1}+2} \leq u, \ldots, X_{k+\tau_{j-1}} > u] \\
&= F^{k-1}(u)(1-F(u)) \qquad (2.139)
\end{aligned}$$

will also be distributed according to a geometric distribution in the i.i.d. case.

In addition, as noted in Devroye & Devroye (1986), we can also state the result in Lemma 2.5

Lemma 2.5 *Given i.i.d. loss data, $\{X_1, X_2, \ldots\}$, from a single risk process with severity distribution $F(x)$, a threshold level u and i.i.d. inter-arrival times between exceedances given by the difference times $\{\tau_1 - 0, \tau_2 - \tau_1, \tau_3 - \tau_2, \ldots\}$, then conditional on having observed a total of T_n exceedances out of the observed losses, the distribution of the time indexes at which the exceedances occurred, is given by*

$$\begin{aligned}
&p(\tau_1, \ldots, \tau_{T_n}|T_n) \\
&\quad = p(\tau_{T_n}|\tau_1, \ldots, \tau_{T_n-1}, T_n)p(\tau_{T_n-1}|\tau_1, \ldots, \tau_{T_n-2}, T_n) \cdots p(\tau_1|T_n). \qquad (2.140)
\end{aligned}$$

In the case in which the number of losses asymptotically grows, i.e. for $n \to \infty$, for a given time interval, say one year, then the random variable $\tau_1|T_n$ is uniform over the times $\{t_1, t_2, \ldots, t_n\}$ where t_i is the time of the i-th loss. The random variable $\tau_2|\tau_1, T_n$ is then uniform over $\{\tau_1, \ldots, t_n\}$ and in general $\tau_i|\tau_{i-1}, T_n$ is uniform over the set $\{\tau_{i-1}, \ldots, t_n\}$.

Next we would like to utilize this information on the distribution of the exceedance times as well as our knowledge of the proposed severity distribution $F(x)$ from which the i.i.d. losses are drawn and the threshold level u to define the distribution for the jth exceedance. Furthermore, we note that because the data is assumed i.i.d., then the exceedances $\{X_{\tau_k}\}_{k \geq 1}$ will also form a set of i.i.d. random variables, each of which will be distributed according to the distribution of exceedance of a loss at a threshold u for a severity distribution F, denoted by $F_{X_{\tau_j} > u}$. Finding this distribution involves invoking the result often referred to as the *Decoupage de Lévy*, see Resnick (2007, chapter 4), and detailed in the context of the single risk OpRisk modeling considered in Lemma 2.6.

Lemma 2.6 (Decoupage de Lévy) *Consider a single risk process in OpRisk in which the losses $\{X_n, n \geq 1\}$ are i.i.d. random variables with distribution function $F(x)$ which has support on $\mathbb{R}^+$. Define on the support of these loss random variables the 'Borel' set $\mathscr{B}$ comprised of countable unions of open sets, countable intersections of open sets and relative complements of open sets in $\mathbb{R}^+$. Then consider a set $B \subseteq \mathscr{B}$ for which $\Pr[X_1 \in B] > 0$. Next, define the index functions $f^+(i)$*

and $f^-(i)$ corresponding to the indexes j for which $X_j \in B$ according to

$$f^+(i) = \begin{cases} 0, & i = 0, \\ \inf\left\{j > f^+(i-1) : X_j \in B\right\}, & i \geq 1, \end{cases}$$
$$f^-(i) = \begin{cases} 0, & i = 0, \\ \inf\left\{j > f^-(i-1) : X_j \in \overline{B}\right\}, & i \geq 1. \end{cases} \tag{2.141}$$

Then defining $N(n) = \sup\{i : f^+(i) \leq n\}$ results in the independence of the sub-set of losses in B denoted by $\{X_{f^+(i)}\}$ and the sub-set of losses in the complement of B denoted by $\{X_{f^-(i)}\}$. In addition, the number of losses for any given i, denoted by the set $\{N(i), i \geq 1\}$ is also independent of both $\{X_{f^+(i)}\}$ and $\{X_{f^-(i)}\}$. Furthermore, $\{X_{f^+(i)}\}$ are i.i.d. with

$$\begin{aligned} \mathbb{P}\mathrm{r}\left[X_{f^+(i)} \in A\right] &= \mathbb{P}\mathrm{r}\left[X_1 \in A | X_1 \in B\right],\ A \subset B, A \in \mathscr{B}, \\ \mathbb{P}\mathrm{r}\left[X_{f^-(i)} \in A\right] &= \mathbb{P}\mathrm{r}\left[X_1 \in A | X_1 \notin B\right],\ A \subset B^c, A \in \mathscr{B}. \end{aligned} \tag{2.142}$$

In addition, $N(n)$ for $n \geq 1$ is a renewal counting process and the mean number of events is given by

$$\mathbb{E}\left[N(n)\right] = n\mathbb{P}\mathrm{r}[X_1 \in B]. \tag{2.143}$$

To understand the use of this result, consider the set B corresponding to the interval (u, ∞), which is the event of losses above the threshold u and assume that the severity distribution $F(x)$ is defined on $\mathbb{R}^+$. Next, consider the exceedance times we defined earlier over the indexes of the times of each loss that we denoted for the jth exceedance by τ_j. Then we see this mapping index $f^+(i)$ as the times (indexes over the times of the losses X_i occurring at t_i) that are produced from the loss time indexes, i.e. the $\tau_1, \tau_2, \ldots$ times of exceedance for which $X_i \in B = (u, \infty)$, that is, the sub-set of losses characterized by $\{X_{f^+(i)}\}$ are the losses which exceeded the threshold from the set of observed losses $\{X_1, X_2, \ldots, X_n\}$ and the corresponding sub-set of losses $\{X_{f^-(i)}\}$ are those losses that failed to exceed the threshold, hence, were in the complement of B, that is, in the range $[0, u]$.

We can now utilize the results above to consider in the i.i.d. case, for a given threshold level u, the distribution of the exceedances given by losses $\{X_{\tau_j}, j \geq 1\}$. This result is stated in Lemma 2.7.

Lemma 2.7 *Consider a set of n losses and times of occurrence in an interval of time $[0, T]$ denoted by $\{(X_i, t_i)\}_{i=1:n}$. Furthermore, consider the set of exceedance times $\{\tau_j\}_{j=1:J}$ for $J \leq n$. As the number of losses $n \to \infty$ and the number of exceedances correspondingly grows, $J \to \infty$, the distribution of the exceedances $\{X_{\tau_j}\}_{j=1:J}$, for a threshold level u on the loss amounts, is given for i.i.d. losses by the joint distribution of*

$$\begin{aligned} F_{X_{\tau_1} > u, X_{\tau_2} > u, \ldots, X_{\tau_J} > u}(\boldsymbol{x} | u, n) &= \mathbb{P}\mathrm{r}[X_{\tau_1} > u, X_{\tau_2} > u, \ldots, X_{\tau_J} > u | u, n] \\ &= \prod_{j=1}^{J} \mathbb{P}\mathrm{r}[X_{\tau_j} > u | u, n] = \left(F_{X_{\tau_1} > u}(x | u, n)\right)^J, \end{aligned} \tag{2.144}$$

where the distribution of the j-th exceedance amount is given by

$$F_{X_{\tau_j}>u} = \mathbb{P}\mathrm{r}[X > x|X > u] = \begin{cases} \dfrac{1-F(x)}{1-F(u)}, & x > u, \\ 1, & x \leq u. \end{cases} \tag{2.145}$$

To understand where this result arises from, consider the first exceedance time and the corresponding loss and its distribution given by $F_{X_{\tau_1}>u}(x)$, as $n \to \infty$ in time interval $[0, T]$. One can decompose this distribution according to the equation

$$\begin{aligned} F_{X_{\tau_1}>u}(x) &= \mathbb{P}\mathrm{r}[X_{\tau_1} > x] = \sum_{k=1}^{\infty} \mathbb{P}\mathrm{r}[\tau_1 = k, X_{\tau_1} > x] = \sum_{k=1}^{\infty} \mathbb{P}\mathrm{r}[\tau_1 = k, X_k > x] \\ &= \sum_{k=1}^{\infty} \mathbb{P}\mathrm{r}[X_1 \leq u, \ldots, X_{k-1} \leq u, X_k > u, X_k > x], \end{aligned} \tag{2.146}$$

where the probability that $x > u$ for the i.i.d. losses is given by $1 - F(u)$ giving

$$\begin{aligned} F_{X_{\tau_1}>u}(x) &= \mathbb{P}\mathrm{r}[X_{\tau_1} > x] = \sum_{k=1}^{\infty} \prod_{s=1}^{k-1} \mathbb{P}\mathrm{r}[X_s \leq u] \mathbb{P}\mathrm{r}[X_k > u, X_k > x] \\ &= \sum_{k=1}^{\infty} (F(u))^{k-1} (1 - F(x|x > u)) \\ &= \sum_{k=1}^{\infty} (F(u))^{k-1} \overline{F}(x|x > u) \\ &= \frac{\overline{F}(x|x > u)}{1 - F(u)} = \frac{\overline{F}(x|x > u)}{\overline{F}(u)}. \end{aligned} \tag{2.147}$$

Hence, we can now proceed to complete the understanding of the Pickands–Balkema–deHaan theorem for the convergence of the distribution of threshold exceedances to the maximal domain of attraction of the GEV family of EVT models. We will consider two illustrative scenarios, one in which the distribution $F(x)$ from which the losses are drawn has a Pareto tail after a threshold level x_* and then the other scenario just assumes properties of the rate of tail decay of the distribution $F(x)$ after a loss threshold x_*.

2.5.1.1 Peaks Over Threshold Pareto Case. Here we consider the case in which one assumes that the severity distribution from which the losses in the single loss process considered above, $\{(X_i, t_i)\}_{i=1:n}$, can be assumed to be drawn from a model in which $F(x|x > x_*)$ is Pareto after a threshold of x_*. This implies that we consider the tails of the severity distribution according to the assumption

$$\overline{F}(x|x > x_*) = \mathbb{P}\mathrm{r}[X > x|X > x_*] = x_*^{\alpha} x^{-\alpha}, \; x > x_*, \; \alpha > 0. \tag{2.148}$$

Next we consider what happens to the distribution of exceedances $\{X_{\tau_j}\}_{j=1:J}$, as the number of losses $n \to \infty$ and the number of exceedances correspondingly grows $J \to \infty$, for

a threshold level on the loss amounts $u \to \infty$, for i.i.d. losses which we found previously had a distribution for the τ_jth exceedance given by

$$F_{X_{\tau_j}>u}(x) = \mathbb{P}\mathrm{r}[X > x | X > u] = \begin{cases} \frac{1-F(x)}{1-F(u)}, & x > u, \\ 1, & x \leq u. \end{cases} \tag{2.149}$$

Under the assumption on the tails of the severity distribution for any threshold level $u > x_*$, one has the distribution of the POT model given explicitly by

$$F_{X_{\tau_j}>x_*}(x|X > x_*) = \frac{\overline{F}(x|X > x_*)}{\overline{F}(x_*)} = \frac{x^{-\alpha}}{x_*^{-\alpha}}, \tag{2.150}$$

which is clearly showing that the distribution of the exceedance for exceedance time τ_j has a Pareto distribution tail after the threshold level exceeds the point x_*, $u > x_*$. In other words, the set of relative exceedances $\{(X_{\tau_j}/x_*), j \geq 1\}$ will be an i.i.d. sample from a Pareto distribution.

2.5.1.2 Peaks Over Threshold Slowly Varying Right Tail Case. Here we consider the case in which one assumes that the severity distribution, from which the losses in the single loss process considered above, $\{(X_i, t_i)\}_{i=1:n}$ can be assumed to be drawn from a model in which $F(x|x > x_*)$, after a threshold of x_* the tails of the severity distribution are regularly varying. This means that they decay according to the assumed relationship

$$\overline{F}(x|x > x_*) = \mathbb{P}\mathrm{r}[X > x | X > x_*] = x^{-\alpha} L(x),\ x > x_*,\ \alpha > 0, \tag{2.151}$$

where $L(x)$ is a slowly varying function of x that satisfies

$$\lim_{t\to\infty} \frac{L(xt)}{L(x)} = 1. \tag{2.152}$$

Again, we consider what happens to the distribution of exceedances $\{X_{\tau_j}\}_{j=1:J}$, as the number of losses $n \to \infty$ and the number of exceedances correspondingly grows $J \to \infty$, for a threshold level on the loss amounts $u \to \infty$, for i.i.d. losses which we found previously had a distribution for the τ_jth exceedance given by

$$F_{X_{\tau_j}>u}(x) = \mathbb{P}\mathrm{r}[X > x | X > u] = \begin{cases} \dfrac{1-F(x)}{1-F(u)}, & x > u, \\ 1, & x \leq u. \end{cases} \tag{2.153}$$

Under the assumption on the tails of the severity distribution for any threshold level $u > x_*$, one has the distribution of the POT model given for the relative exceedances by

$$\mathbb{P}\mathrm{r}\left[\frac{X_{\tau_1}}{x_*} > x\right] = \frac{\overline{F}(x_* x)}{\overline{F}(x_*)} \approx x^{-\alpha},\ x > 0,\ x_* \text{large}, \tag{2.154}$$

which shows that the distribution of the exceedance for exceedance time τ_j has an approximate Pareto distribution tail after the threshold level exceeds the point x_*, $u > x_*$. In other words, the set of relative exceedances $\{(X_{\tau_j}/x_*), j \geq 1\}$ will be an i.i.d. sample approximately from a Pareto distribution, such that this approximation becomes exact asymptotically as $x_* \to \infty$.

2.5.1.3 Peaks Over Threshold via Domain of Attraction Considerations and the GPD. Understanding the POT convergence result specified according to the Pickands–Bulkema–deHaan results in Theorem 2.7 can also be achieved through the consideration of the domain of attraction conditions presented for the block maxima discussions. In particular, we recall Condition D_{II} in Table 2.1 specified according to the following convergence condition on the loss distribution $F(x)$ to be in the domain of attraction of the GEV family $F \in \mathscr{D}(M)$, given by

$$\frac{1 - F(y + b(y)v)}{1 - F(y)} \to (1 + \gamma v)^{-1/\gamma}, \tag{2.155}$$

as $y \to x_* = \inf\{x : F(x) = 1\}$.

There are generally two approaches to consideration of the POT's model either from the point process perspective discussed earlier or directly from the domain of attraction conditions.

An analogous approach to the results on the distribution of threshold exceedances discussed earlier is to consider the domain of attraction conditions and the GPD in the case in which one considers a point process approximation, where one assumes generation of a loss process with severity $F \in \mathscr{D}(M)$ for which an approximate model for the exceedances over a high threshold is given by a GPD. That is, the severity model, $F(x)$, is a member of the Frechet sub-family of the GEV distributions which satisfied Condition D_{II}.

In particular, one considers the loss process $\{(X_i, t_i)\}_{i=1:n}$ as above to define a marked point process in which the times $\{t_i : i \geq 0\}$ form a point process, and the marks correspond to random variables $\widetilde{X}_i$ such that

$$\widetilde{X}_i = \begin{cases} X_i, & x \geq u \text{ exceedances of threshold } u, \\ 0, & x < u, \end{cases} \tag{2.156}$$

giving the marked point process defined by counts for time interval $A = [0, T]$ given by

$$\begin{aligned} N(A) &= \sum_{i=1:n} \widetilde{X}_i \mathbb{I}(t_i \in A) \\ &= \sum_{\widetilde{\tau}_j = 1:T_n} X_{\widetilde{\tau}_j} \mathbb{I}(\widetilde{\tau}_j \in A), \end{aligned} \tag{2.157}$$

which we see is equivalent to the process defined over the T_n exceedance times. It is also possible to consider an equivalent representation of this marked counting process or, in general, renewal process as a bivariate counting process by extending the state space. Here we consider such a representation defined by the two-dimensional point process on the product space $(0, T) \times (u, \infty)$ given by the tuple defined over the scaled losses X_i and their indexes by

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i \in \{1, 2, \ldots, n\} \right\}, \tag{2.158}$$

where the values a_n and b_n are selected to normalize the maximum loss $X_{(n)}$ as detailed in the previous discussions on block maxima. A representation of such a bivariate counting process is provided in Figure 2.19, which builds on the representation of the POT data in Figure 2.18. In this case, if one considers sets A which exclude the lower boundary, then the distribution of P_n will converge weakly to a two-dimensional Poisson processes. The intensity measure

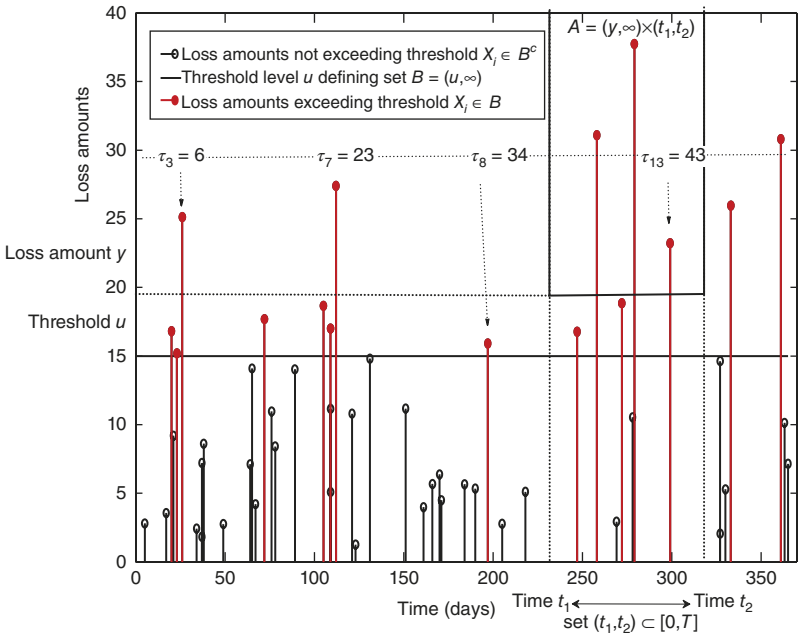

FIGURE 2.19 **An example realization of a single risk process $\{(X_i, t_i)\}_{i=1:n}$ with marked losses that exceeded a threshold loss amount u. The times on the x-axis at which the losses occurred correspond to the days t_i, the amounts correspond to the values X_i and the secondary process comprised of a subset of the losses, denoted by τ_j's as illustrated on the figure, correspond to the indexes of the jth exceedance. In addition, an example region $A = (t_1, t_2) \times (y, \infty)$ is marked for points exceeding threshold u in time interval $(t_1, t_2) \subset [0, T]$. *See insert for color representation of this figure.***

of this process is given by considering the limiting distribution of the process, as $n \to \infty$, under the condition that no points are in the interval $(0, 1)$. That is, one considers for all $i \in \{1, 2, \ldots, n\}$ the limit of the probability of the bivariate process not having a point in the set $(0, 1) \times (x, \infty)$, which can be re-expressed in terms of a probability of the standardized maximum in the limit of the number of losses which we know under the assumption that $F \in \mathscr{D}(M)$ will be analytic and is given by

$$\lim_{n\to\infty} \Pr\left[\left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n}\right) \notin (0,1)\times(x,\infty)\right] = \lim_{n\to\infty} \Pr\left[\left(\frac{X_{(n)} - b_n}{a_n}\right) < x\right]$$
$$= \exp\left(-(1+\gamma x)^{-1/\gamma}\right). \quad (2.159)$$

Returning to the specification of the resulting intensity measure, one would obtain for subsets $A = (t_1, t_2) \times (x, \infty)$, such that $t_1 < t_2$ the function given in Beirlant *et al.* (2004, equation 5.14)

$$\Lambda(A) = \mathbb{E}[N(A)] = (t_1 - t_2)(1 + \gamma x)^{-\frac{1}{\gamma}}. \quad (2.160)$$

Finally, considering the threshold $u \to \infty$, one has the probability of exceedance for the standardized ith loss that is approximately given by the Poisson probability

$$\begin{aligned}
&\mathbb{P}\mathrm{r}\left[\frac{X_i - b_n}{a_n} > u + x \middle| \frac{X_i - b_n}{a_n} > u\right] \\
&\quad = \frac{\mathbb{P}\mathrm{r}\left[\left(X_i - b_n\right)/a_n > u + x\right]}{\mathbb{P}\mathrm{r}\left[\left(X_i - b_n\right)/a_n > u\right]} \approx \frac{\Lambda((0,1) \times (u + x, \infty))}{\Lambda((0,1) \times (u, \infty))} \\
&\quad = \left(1 + \frac{\gamma x}{1 + \gamma u}\right)^{-1/\gamma},
\end{aligned} \tag{2.161}$$

which we see recovers the GPD survival function with scale given by $\sigma(u) = 1 + \gamma u$.

Another approach one may adopt to utilize the domain of attraction condition is to consider the distributional ratio for the severity distribution of the losses in the single risk process corresponding to $(1 - F(y + b(y)v))/(1 - F(y))$, as discussed in Beirlant *et al.* (2004, section 5.3.1); this ratio can be re-interpreted as the conditional survival function of the exceedances $Y = X - u$ over a threshold u, taken at $yb(u) > 0$ according to the equation

$$\begin{aligned}
\overline{F}_{X>u}(x) &= 1 - F_{X>u}(x) \\
&= \mathbb{P}\mathrm{r}[X - u \leq x | X > u],\ 0 \leq x \leq x_* - u \\
&:= \mathbb{P}\mathrm{r}[Y > yb(u) | Y > 0] \\
&= \mathbb{P}\mathrm{r}\left[\frac{X - u}{b(u)} > y | X > u\right] \\
&= \frac{1 - F(u + b(u)y)}{1 - F(u)}.
\end{aligned} \tag{2.162}$$

Therefore, we see from Condition D_{II} that a natural estimation of this exceedances distribution is to consider the approximation in the tails given by the Frechet sub-family member, the GPD, in which the scale parameter is given by $\sigma = b(u)$,

$$\overline{F}_{X>u}(x) \sim \left(1 + \frac{\gamma y}{b(u)}\right)^{-1/\gamma}. \tag{2.163}$$

The POT's approach to modeling in OpRisk is particularly important in practice to address questions related to the design of reporting mechanisms for OpRisk loss processes. It is a common practice to specify a reporting threshold below which losses in a single risk processes remain unreported. This form of censoring of the data is often not based on the statistical justification but more on practical business environment and computing IT justifications. It is clear that a statistical model under the simplifying assumptions of i.i.d. data from a fixed severity model can guide in an understanding of the expected number of observed losses for a given threshold in a given period of time. It can, therefore, be directly utilized to guide in an informed setting of reporting thresholds for loss databases.

2.6 Estimation Under the Peaks Over Threshold Approach via the Generalized Pareto Distribution

Example LDA models are presented for the GPD model in the Frechet family. Here we focus on the aspects of parameter estimation under such a severity model. As with the GEV modeling

approach, there are also numerous approaches to estimation of the GPD model parameters for the severity in the LDA framework. We consider several proposed statistical estimation approaches and discuss their merits in the context of OpRisk.

The number of estimation procedures discussed in the literature for fitting the GPD model, primarily results from two statistical considerations which have practical consequences for estimation of such models:

1. firstly, the implications of parameter restrictions and existence of solutions under particular estimation procedures given these restrictions; and
2. secondly, the behaviour of the parameter estimates in terms of bias and variance and how robust such estimators are if data is not actually drawn exactly from the GPD model. In other words, the behaviour of the estimations if the statistical assumptions required for the threshold exceedances theory and the resulting GPD model to hold are violated in some respect.

The estimation approaches discussed in the literature include robust and non-robust estimation methods given by

1. MLE under the GPD model, see Smith (1985), Grimshaw (1993);
2. MOM, probability-weighted moments and generalizations, see Hosking & Wallis (1987);
3. Elemental Percentile Method; see Castillo & Hadi (1997);
4. Minimum density power divergence estimators (MDPDE), see Basu *et al.* (1998) and Juárez & Schucany (2004);
5. Method of medians, see Peng & Welsh (2001);
6. Optimally biased robust estimator (OBRE), see Dupuis (1999);
7. Principle of maximum entropy, see Singh & Guo (1995).

First, we discuss how one may go about preparing the data for the POT method in practical OpRisk settings.

Assume that for given threshold value u and the observations from the single risk process $\{X_i\}_{i=1:n}$, there were J exceedances under the assumption that $X_i \sim F(x)$. The data for which $X_i > u$ is denoted by the sub-set $\widetilde{\boldsymbol{X}} = \{X_{\tau_1}, X_{\tau_2}, \ldots, X_{\tau_J}\}$.

Algorithm 2.10 Preparing the Data Under a POT's Approach

1. *Take the full set of N losses observed in a period $[0, T]$, time ordered and denoted by $\{X_1(t_1), \ldots, X_N(t_N)\}$ for $0 \leq t_1 < t_2 < \cdots < t_N \leq T$, where the ith loss occurred at time t_i. Given a threshold u, typically set in practice according to an order statistic $X_{(N-J)}$ such that $u = X_{(N-J)}$ to ensure enough data to perform estimation. In this case, there will be exactly J exceedances of the threshold.*
2. *Extract the losses that exceed the threshold $\{X_{\tau_j} : X_{\tau_j} > u\}_{j=1:J}$, which could be, for example, if $u = X_{(N-J)}$ the set of order statistics and their associated exceedance times denoted by $\{(X_{(N-j)}, t_{\tau_j})\}_{j=0:J}$.*
3. *Obtain the absolute exceedances given by the set $\{Y_j\}_{j=1:J}$ such that*

$$Y_j = X_{(N-j)} - u = X_{(N-j)} - X_{(N-J)}$$

with $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_J)$.

Next we present a brief discussion on some of the different approaches statisticians have adopted on estimation of the GPD model parameters. We note that, typically, when performing estimation of the GPD model under a POT's methodology in OpRisk, we would have a limited amount of data and one would typically consider the extreme risk settings to have small-sample-size characteristics. Hence, we focus in the following discussions on the topic of parameter estimation in the POT's GPD exceedances model in situations in which the sample size is sufficiently small for several convenient large sample asymptotic results to not hold.

Practical Guides to Estimations of GPD Parameters. The following specification guides due to Hosking & Wallis (1987) and Castillo & Hadi (1997) have been provided in the literature as a guide to the different estimation techniques available for fitting GPD models to data.

1. In settings in which one believes that the data was not from a heavy-tailed severity model, not the models considered in this chapter, it is a common specification to restrict to cases in which $-1/2 < \gamma < 1/2$. In this range, it was proposed in Hosking & Wallis (1987) that the following guides can be applied:
   (a) In general, the large sample asymptotics that provide the asymptotic efficiency for the maximum-likelihood estimators will not apply unless the sample size exceeds $n > 500$. In addition, the MLE estimates do not exist for ranges $\gamma > 1$, which is an interesting range suitable for heavy-tailed models.
   (b) The MOM-based approach to GPD parameter estimation is unreliable for ranges of $\gamma < -0.2$.
   (c) The method of PWM is reliable for ranges $-1/2 < \gamma < 0$.
2. Particular interest in OpRisk settings in which losses come from a heavy-tailed severity model would require consideration of infinite mean and infinite variance specifications. This requires $\gamma > 1/2$. As noted in examples in Castillo & Hadi (1997) in which $\gamma \notin (-0.5, 0.5)$ typically arise in settings with heavy-tailed models and truncated models, both of which are of direct relevance to OpRisk modeling. In such cases, they recommend the application of the elemental percentile method (EPM).
3. The modified version of the MOM denoted as the PWM-based approach of Hosking & Wallis (1987) was developed for the GPD model with the shape (EVI) in the range $\gamma < 1$. It should also be noted that when $\gamma > 1$, the second and higher moments do not exist and, hence, the MOM and PWM estimates do not exist. In Castillo & Hadi (1997), it is argued that MOM and PWM approaches may produce estimates which are inconsistent with observed data, and they proposed the EPM, which are defined over all values of γ.
4. The MDPDE approach is proposed if samples are 'contaminated' or may not come directly from a GPD model but instead from some alternative process for which the GPD is being used to approximate the severity distribution. In such cases, Juárez & Schucany (2004) propose the MDPDE in place of the MLE estimator as they are more statistically efficient. In addition, it is claimed that the medians based estimator performed poorly.

2.6.1 MAXIMUM-LIKELIHOOD ESTIMATION UNDER THE GPD MODEL

Here we recall one of the examples provided in the GEV-LDA sub-family of models in the Frechet–Pareto family, namely, the subfamily of the generalized Pareto. Under this model, one

considered a random variable $X \sim GPD(\gamma, \sigma)$ to have a density given by

$$f_X(x; \gamma, \sigma) = \frac{\sigma^{1/\gamma}}{(\sigma + \gamma x)^{(1/\gamma)+1}} \tag{2.164}$$

with shape parameter γ and scale parameter σ. Then under the assumptions detailed earlier, for i.i.d. exceedances in a POT's approach, the log-likelihood for the GPD as a function of the absolute exceedance data is given, for the case in which $1 + (\gamma Y_j/\sigma) > 0$, by

$$\ell(\boldsymbol{Y}; \gamma, \sigma) = -J \ln \sigma - \left(\tfrac{1}{\gamma} + 1\right) \sum_{j=1}^{J} \ln\left(1 + \left(\gamma Y_j/\sigma\right)\right), \tag{2.165}$$

where the condition $1 + (\gamma Y_j/\sigma) > 0$ ensures that the log-likelihood is finite.

If $\gamma = 0$, the log-likelihood is given according to the exponential-based distribution according to the equation

$$\ell(\boldsymbol{Y}; 0, \sigma) = -J \ln \sigma - \tfrac{1}{\sigma} \sum_{j=1}^{J} Y_j. \tag{2.166}$$

Given the loglikelihood, the moments of the GPD or the quantile function, there are numerous statistical approaches one could adopt to perform the parameter estimation. First, we discuss how to perform MLE for such models. Maximization of the GPD likelihood provided in Equations 2.165 and 2.166 with respect to the parameters γ and σ is subject to the constraints:

1. $\sigma > 0$

2. $1 + \gamma y_{(J)}/\sigma > 0$ where $y_{(J)} = \max\{y_1, y_2, \ldots, y_J\}$.

This second constraint is important because one observes that if $\gamma < -1$ as $-\sigma/\gamma \to y_{(J)}$, then the likelihood approaches infinity. Hence, to obtain maximum-likelihood parameter estimates, one should maximize the likelihood subject to these constraints and $\gamma \geq -1$.

Reparameterization of the GPD Log-Likelihood and Maximization. In practice, it is beneficial to consider a reparameterizations of the GPD log-likelihood function according to the equation

$$(\gamma, \sigma) \to (\gamma, \tau) = \left(\gamma, \frac{\gamma}{\sigma}\right), \tag{2.167}$$

producing a reparameterized log-likelihood model given by

$$\ell(\boldsymbol{Y}; \gamma, \tau) = -J \ln \gamma + J \ln \tau - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{J} \ln(1 + \tau Y_i). \tag{2.168}$$

This log-likelihood is then maximized subject to $\tau < 1/y_{(J)}$ and $\gamma \geq -1$ giving under the first-order partial derivative

$$\frac{\partial \ell(\boldsymbol{Y}; \gamma, \tau)}{\partial \gamma} = 0;$$

$$\Rightarrow \ \gamma(\tau) = \frac{1}{J} \sum_{j=1}^{J} \ln(1 - \tau y_j). \tag{2.169}$$

Hence, the estimation is performed in two steps as follows:

1. Estimate $\widehat{\tau}^{MLE} = \arg\max \ell(\gamma(\tau), \tau)$ subject to $\tau < 1/y_{(J)}$.
2. Estimate $\widehat{\gamma}^{MLE} = \frac{1}{J}\sum_{j=1}^{J} \ln(1 - \widehat{\tau}^{MLE} y_j)$. Then solve for the original parameterizations via inversion $\widehat{\sigma}^{MLE} = -\widehat{\gamma}^{MLE}/\widehat{\tau}^{MLE}$.

Note that the log-likelihood $\ell(\gamma(\tau), \tau)$ is continuous at $\tau = 0$, hence is the estimator $\widehat{\tau}^{MLE} = 0$, then one should consider $\widehat{\gamma}^{MLE} = 0$ and

$$\widehat{\sigma}^{MLE} = \frac{1}{J}\sum_{j=1}^{J} y_j. \tag{2.170}$$

In addition, in practice, to ensure that $\gamma \geq 1$, the condition that $\tau < 1/y_{(J)}$ should be modified to $\tau < (1-\varepsilon)/y_{(J)}$, where ε is found from the condition that $\gamma(\tau) \geq -1$.

The covariances of the MLE parameters can be estimated using the inverse of the 'observed Fisher information matrix' given for J realized data samples $\boldsymbol{y} = [y_1, y_2, \ldots, y_J]$ for the (k, m)-th element of the covariance by

$$\widehat{\mathscr{F}}(\widehat{\boldsymbol{\theta}})_{k,m} = -\frac{1}{J}\sum_{j=1}^{J} \left.\frac{\partial^2 \ell(y_j|\,\boldsymbol{\theta})}{\partial\theta_k \partial\theta_m}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}. \tag{2.171}$$

In the case of the GPD MLE parameter estimation, the required second-order partial derivatives are given by

$$\begin{cases} \dfrac{\partial^2}{\partial\gamma^2}\ell(\boldsymbol{y};\gamma,\sigma) = -\dfrac{2}{\gamma}\displaystyle\sum_{j=1}^{J}\frac{y_j}{\sigma+\gamma y_j} + \left(1+\frac{1}{\gamma}\right)\sum_{j=1}^{J}\frac{y_j^2}{(\sigma+\gamma y_j)^2} \\ \dfrac{\partial^2}{\partial\sigma^2}\ell(\boldsymbol{y};\gamma,\sigma) = -\dfrac{1}{\sigma}(1+\gamma)\displaystyle\sum_{j=1}^{J}\frac{y_j}{(\sigma+\gamma y_j)^2} \\ \dfrac{\partial^2}{\partial\sigma\partial\gamma}\ell(\boldsymbol{y};\gamma,\sigma) = \displaystyle\sum_{j=1}^{J}\frac{y_j}{(\sigma+\gamma y_j)^2} - \frac{1}{\sigma}\sum_{j=1}^{J}\frac{y_j^2}{(\sigma+\gamma y_j)^2}. \end{cases} \tag{2.172}$$

Remark 2.16 *It has been shown in Smith (1985) and (Embrechts et al. 1997, section 6.5.1) that in the case in which $\gamma > -1/2$, the MLE vector $\widehat{\boldsymbol{\theta}}^{MLE} = (\widehat{\gamma}^{MLE}, \widehat{\sigma}^{MLE})$ is asymptotically consistent and distributed according to a bivariate Gaussian distribution with asymptotic covariance obtained using the MLE parameter estimates and the inverse Fisher information matrix. The Fisher information matrix is given analytically for the (k, m)th element as*

$$\begin{aligned} \mathscr{I}(\boldsymbol{\theta})_{k,m} &= \mathbb{E}\left[\frac{\partial}{\partial\theta_k}\ln f(Y_1|\,\boldsymbol{\theta})\frac{\partial}{\partial\theta_m}\ln f(Y_1|\,\boldsymbol{\theta})\right] \\ &= -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_k\partial\theta_m}\ln f(Y_1|\,\boldsymbol{\theta})\right], \end{aligned} \tag{2.173}$$

where the limit is asymptotically given by

$$\lim_{J\to\infty}\left[\mathscr{I}(\boldsymbol{\theta})\right]^{-1}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{MLE}} = \frac{1}{J}\left.\begin{bmatrix} (1+\gamma)^2 & -(1+\gamma)\sigma \\ -(1+\gamma)\sigma & 2(1+\gamma)\sigma^2 \end{bmatrix}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{MLE}}. \tag{2.174}$$

2.6.1.1 Small Sample Considerations for MLE Estimation in GPD. It is well documented in the statistical literature that the estimation of the parameters in the GPD family can run into difficulties and may not work well in small samples, see Castillo & Hadi (1997) and Hosking & Wallis (1987). This is largely why so many other approaches have been developed for such methods in the statistical literature. In practical settings of OpRisk, this is a concern, because, typically, the limited number of large losses exceeding a given threshold will be small in number, resulting in this concern warranting more consideration. In this section, we detail a technical discussion about what exactly happens to GPD likelihood in small sample settings.

It is noted that the MLE of the GPD in large sample asymptotic cases is discussed in Davison (1984) and Smith (1985). If one considers the parameterizations of the GPD density for a random variable $X \sim GPD(\gamma, \widetilde{\sigma})$ given by

$$f_X(x; \gamma, \widetilde{\sigma}) = \frac{1}{\widetilde{\sigma}|\gamma|}\Big(1 + s\frac{x}{\widetilde{\sigma}}\Big)^{-(1+\gamma)/\gamma}, \tag{2.175}$$

where $\widetilde{\sigma} = \sigma/|\gamma|$, $s = \text{sign}(\gamma)$ and when $\gamma < 0$, the domain becomes $0 < x < \sigma$. Hence, one can combine in this parameterization the two sub-families discussed earlier for $\gamma < 0$ and $\gamma > 0$.

Under this parameterization, the log-likelihood divided by the sample size n is given by

$$\ell(\mathbf{y}; \gamma, \widetilde{\sigma}) = -\ln(s\gamma\widetilde{\sigma}) - \frac{1+\gamma}{\gamma n}\sum_{i=1}^{n}\ln\Big(1 + s\frac{y_i}{\widetilde{\sigma}}\Big) \tag{2.176}$$

and it is assumed that if $\gamma < 0$, then $\widetilde{\sigma} > \max\{y_i\}_{i=1}^n$, otherwise the likelihood is zero.

Under this setting, it is well established that the MLE in large sample settings exists and is asymptotically Gaussian and efficient when $\gamma > -0.5$. However, in small samples, anomalous behaviour of the likelihood surface can be encountered for the GPD. In del Castillo & Daoudi (2009), several results are derived for the properties of the small sample GPD likelihood and profile likelihoods with respect to the coefficient of variation, given in Definition 2.29.

Definition 2.29 *The coefficient of variation for a random variable X with finite mean and variance, defined on positive support and taking values on a ratio scale is given by*

$$Vco(X) = \frac{\sqrt{\mathbb{V}\text{ar}(X)}}{\mathbb{E}[X]}. \tag{2.177}$$

The coefficient of variation for the GPD can be found in closed form in terms of the parameters of the distribution.

Remark 2.17 *The coefficient of variation of a random variable with $X \sim GPD(\gamma, \widehat{\sigma})$ when $\gamma < 0.5$ is given by*

$$Vco(X) = \zeta = \sqrt{\frac{1}{1-2\gamma}} > 1. \tag{2.178}$$

However, it is observed in practice that for small sample sizes, the empirically estimated coefficient of variation is given by

$$\widehat{Vco}(X) = \widehat{\zeta} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}{\frac{1}{n}\sum_{i=1}^{n} x_i}}, \tag{2.179}$$

and can be lower than 1.

To understand this, we consider in the GPD model the profile likelihood, obtained by differentiating the log-likelihood in Equation 2.176 with respect to the EVI parameter γ and equating to zero. After some algebra, one obtains the expression in terms of the modified scale parameter $\widetilde{\sigma}$, given by

$$\widehat{\gamma} = \gamma(\widetilde{\sigma}, s) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(1 + s\frac{y_i}{\sigma}\right). \tag{2.180}$$

Therefore, one can state the profile likelihood in the GPD model according to Proposition 2.6.

Proposition 2.6 (Profile Likelihood for GPD) *Given a set of i.i.d. loss realizations* $\boldsymbol{y} = \{y_1, y_2, \ldots, y_n\}$ *from a GPD model, then the resulting profile likelihood* $l_p(\widetilde{\sigma}, s)$ *of the GPD is given by*

$$l_p(\widetilde{\sigma}, s) = -\ln(s\gamma(\widetilde{\sigma}, s)\widetilde{\sigma}) - \gamma(\widetilde{\sigma}, s) - 1. \tag{2.181}$$

Having defined the profile likelihood, the following properties of this expression were observed in del Castillo & Daoudi (2009, propositions 1 and 2, and theorem 1).

Lemma 2.8 *Given the profile likelihood* $l_p(\widetilde{\sigma}, s)$ *for the GPD model in Equation 2.181, then the following limit as the modified scale goes to infinity* $\widetilde{\sigma} \to \infty$ *exists and is expressed in terms of the sample mean according to*

$$l_0 = \lim_{\widetilde{\sigma}\to\infty} l_p(\widetilde{\sigma}, s) = -\ln\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - 1. \tag{2.182}$$

This limiting profile likelihood for the GPD model as the modified scale tends to infinity corresponds to the log-likelihood of the exponential distribution divided by the sample size n.

As a consequence of this result for the limit of the profile likelihood of the GPD family, as the modified scale parameter goes to infinity, one can state the following characterization of the GPD likelihoods behaviour in terms of the estimated sample coefficient of variation $\widehat{\zeta}$ given in Theorem 2.8.

Theorem 2.8 (Characterizing the $l_p(\widetilde{\sigma}, s)$ of GPD by $\widehat{\zeta}$) *One can characterize the behaviour of the MLE of the GPD for small samples according to the two ranges of* $\gamma > 0$ *and* $\gamma < 0$ *as follows.*

1. *Consider the Pareto distribution with shape parameter in the range* $\gamma > 0$*, then the behaviour of the profile likelihood* $l_p(\widetilde{\sigma}, s)$ *of the GPD given by*

$$l_p(\widetilde{\sigma}, s) = -\ln(s\gamma(\widetilde{\sigma}, s)\widetilde{\sigma}) - \gamma(\widetilde{\sigma}, s) - 1 \tag{2.183}$$

is characterized by the estimated coefficient of variation as follows:

- *If* $\widehat{\zeta} > 1$*, then the profile likelihood* $l_p(\widetilde{\sigma}, 1)$ *is a monotonic decreasing function for sufficiently large* $\widetilde{\sigma}$*.*
- *If* $\widehat{\zeta} < 1$*, then the profile likelihood* $l_p(\widetilde{\sigma}, 1)$ *is a monotonic increasing function for sufficiently large* $\widetilde{\sigma}$*.*

2. *Consider the Pareto distribution with shape parameter in the range* $\gamma < 0$*, then the behaviour of the profile likelihood* $l_p(\widetilde{\sigma}, s)$ *of the GPD with bounded support is characterized by the estimated coefficient of variation as follows:*

- *If* $\widehat{\zeta} > 1$*, then the profile likelihood* $l_p(\widetilde{\sigma}, -1)$ *is a monotonic increasing function for sufficiently large* $\widetilde{\sigma}$*.*
- *If* $\widehat{\zeta} < 1$*, then the profile likelihood* $l_p(\widetilde{\sigma}, -1)$ *is a monotonic decreasing function for sufficiently large* $\widetilde{\sigma}$*.*

Finally, it can be shown that when the estimated coefficient of variation is $\widehat{\zeta} > 1$*, then*

$$l_p(\widetilde{\sigma}, 1) > l_0 > l_p(\widetilde{\sigma}, -1), \tag{2.184}$$

and vice versa, if $\widehat{\zeta} < 1$*, then the inequality switches direction to give*

$$l_p(\widetilde{\sigma}, 1) < l_0 < l_p(\widetilde{\sigma}, -1), \tag{2.185}$$

for sufficiently large $\widetilde{\sigma}$*.*

Next we illustrate the profile likelihood functions as contour plots for each of the four cases detailed for the coefficient of variation in Example 2.5.

EXAMPLE 2.5

We consider the four cases for the coefficient of variation given by the combinations of EVI (finite variance case) and estimated coefficient of variation $(\gamma, \widehat{\zeta}) \in \{(0.2, \widehat{\zeta} > 1), (0.2, \widehat{\zeta} < 1), (-0.2, \widehat{\zeta} > 1), (-0.2, \widehat{\zeta} < 1)\}$. This involves generating four sets of i.i.d. observations from the GPD model with these characteristics, each with sample size $n = 5$. The realized data samples satisfying these two conditions are given by

$$(\gamma > 0, \widehat{\zeta} > 1) = (0.2, 19.3), \text{for data}\{7.6531, 9.09, 13.03, 39.09, 9.92\};$$

$$(\gamma > 0, \widehat{\zeta} < 1) = (0.2, 0.69), \text{for data}\{0.64, 2.31, 0.70, 0.04, 0.71\};$$

$$(\gamma < 0, \widehat{\zeta} > 1) = (-0.2, 12.2), \text{for data}\{2.23, 13.26, 3.07, 12.45, 21.36\};$$

$$(\gamma < 0, \widehat{\zeta} < 1) = (-0.2, 0.80), \text{for data}\{1.07, 2.41, 0.93, 0.22, 0.02\}.$$

Given these samples, the kernel density estimate plots of the profile likelihood versus modified scale parameter are provided for each of the four scenarios in Figure 2.20.

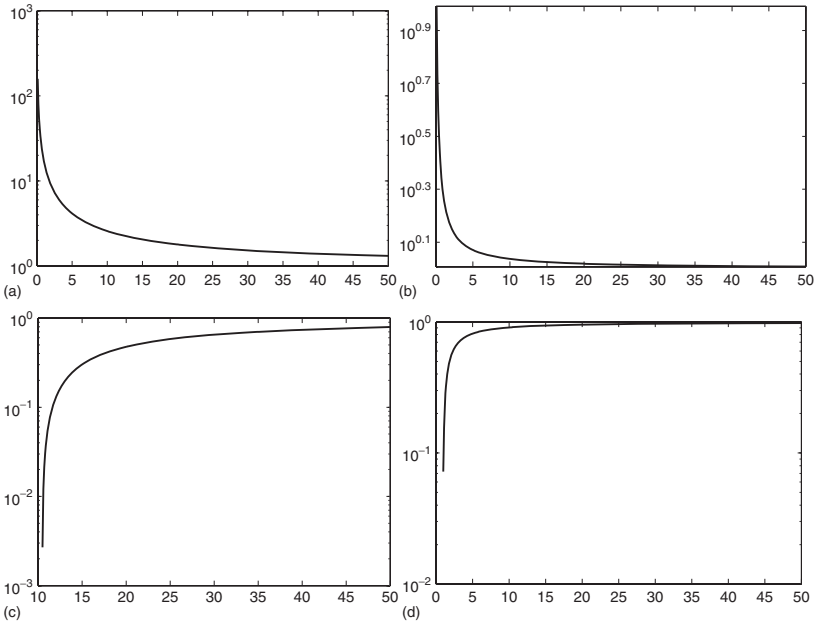

FIGURE 2.20 **Generalized Pareto distribution small sample behaviour of the profile likelihood characterized as a function of the coefficient of variation. (a) Case 1** $(\gamma > 0, \widehat{\zeta} < 1)$**, (b) Case 2** $(\gamma > 0, \widehat{\zeta} < 1)$**, (c) Case 3** $(\gamma < 0, \widehat{\zeta} > 1)$ **and (d) Case 4** $(\gamma < 0, \widehat{\zeta} > 1)$**.**

Remark 2.18 *The implications of the results in Theorem 2.8 are that as* $\widetilde{\sigma} \to \infty$ *then* $|\gamma(\widetilde{\sigma}, s)| \to 0$*. Hence, for* γ *in a neighbourhood of zero, we see that when* $\widehat{\zeta} > 1$*, the Pareto distribution is more likely than the exponential distribution, and if* $\widehat{\zeta} < 1$*, then GPD with a bounded support is more likely than the exponential distribution. These have consequences about the use of the exponential distribution in the MLE of the GPD at small sample sizes.*

2.6.2 COMMENTS ON PROBABILITY-WEIGHTED METHOD OF MOMENTS ESTIMATION UNDER THE GPD MODEL

The approach of moment matching in the GPD model was first proposed in Hosking & Wallis (1987) and utilizes the fact that expressions for the moments of the GPD are known and exist for the rth moment when $\gamma < 1/r$ as non-linear functions of the GPD shape and scale parameters. It has been shown that, given they exist, the mean, variance, skew and kurtosis of a random variable with GPD are given by

$$\begin{aligned}
\mathbb{E}[X] &= \frac{\sigma}{1-\gamma}, \quad \mathbb{V}\text{ar}(X) = \frac{\sigma^2}{(1-\gamma)^2(1-2\gamma)}, \\
\mathbb{S}\text{kewness}(X) &= \frac{2(1+\gamma)(l-2\gamma)^{1/2}}{(1-3\gamma)}, \\
\mathbb{K}\text{urtosis}(X) &= \frac{3(1-2\gamma)(3+\gamma-2\gamma^2)}{(1-3\gamma)(l-4\gamma)} - 3.
\end{aligned} \tag{2.186}$$

Hence, the estimators of the shape and scale parameters obtained from moment matching (MOM) are given according to Definition 2.30.

Definition 2.30 (Method of Moments GPD Model) *The estimators for the shape and scale parameters in the GPD severity distribution by the MOM are given, for observed realized losses prepared under the POT's exceedances method and denoted as* $\{x_i\}_{i=1}^n$*, by the estimators*

$$\begin{aligned}
\widehat{\gamma}^{MOM} &= \frac{1}{2}\left(1 - \frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\left(\frac{1}{n-1}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j\right)^2\right)}\right), \\
\widehat{\sigma}^{MOM} &= \frac{1}{2n}\sum_{i=1}^n x_i \left(1 + \frac{\left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2}{\left(\frac{1}{n-1}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j\right)^2\right)}\right).
\end{aligned} \tag{2.187}$$

This method of estimation has been extended to probability-weighted versions of moment matching utilizing the definition of the PWM of Greenwood *et al.* (1979) of a random variable X given in Definition 2.31. It should be noted that the PWM can also be used to define the L-moments that were considered in estimation of the GEV model parameters.

Definition 2.31 (Probability Weighted Moments) *Consider a univariate continuous random variable* $X \sim F(x)$*. The PWM are defined according to*

$$M_{p,r,s}(X) = \mathbb{E}[X^p F(x)^r (1 - F(x))^s], \tag{2.188}$$

for real-valued powers p, r, s*. Note, when* $r = s = 0$*, one recovers the conventional moments.*

Originally, the motivations provided for consideration of PWM's versus standard moments of a distribution in a moment-matching procedure is that Greenwood *et al.* (1979) demonstrated that many distributions admit simpler expressions in terms of the distributional parameters for the PWM's than for the standard moments. Therefore, solving the systems of non-linear coupled equations in the MOM is simplified under such an approach.

In the case of the GPD model with a random variable $X \sim GPD(\gamma, \sigma)$, the versions of the PWM expressions provided in Hosking & Wallis (1987) are given for the cases $M_{1,0,s}(X)$ in the range $\gamma < 1$ by

$$M_{1,0,1}(X) = \frac{\sigma}{(s+1)(s+1-\gamma)}, \ \gamma < 1. \tag{2.189}$$

In addition, these can be estimated using samples for the cases of $s \in \{0, 1\}$ according to the following estimators,

$$\widehat{M}_{1,0,s}(X) = \frac{1}{n}\sum_{j=1}^n \left(1 - \frac{j}{n+1}\right)^s X_{(j)}, \tag{2.190}$$

where $X_{(j)}$ is the jth order statistic of a sample of size n. This allows one to match the sample-estimated PWM with the analytic expressions to obtain parameter estimators for the shape and scale parameters of the GPD model given for $s \in \{0, 1\}$ by

$$\begin{aligned}\widehat{\gamma}^{PWM(1,0,s)} &= 2 - \frac{\widehat{M}_{1,0,0}(X)}{\widehat{M}_{1,0,0}(X) - 2\widehat{M}_{1,0,1}(X)},\\ \widehat{\sigma}^{PWM(1,0,s)} &= \frac{2\widehat{M}_{1,0,0}(X)\widehat{M}_{1,0,1}(X)}{\widehat{M}_{1,0,0}(X) - \widehat{M}_{1,0,1}(X)}.\end{aligned} \tag{2.191}$$

Remark 2.19 *The two approaches of MOM and PWM should be utilized with caution, because as pointed out in Castillo & Hadi (1997), for a given set of observed losses for which one is fitting the GPD model too, the resulting MOM or PWM estimates of the parameters may be inconsistent with the observations.*

2.6.3 ROBUST ESTIMATORS OF THE GPD MODEL PARAMETERS

In this section, we detail an additional parameter estimation procedure which may be developed with no restriction on the EVI shape parameters $\gamma \in \mathbb{R}$. This is based on the percentile-based matching approach proposed in Castillo & Hadi (1997).

2.6.3.1 Elemental Percentile Method and the Pickands Estimator. The idea of the EPM approach to parameter estimation involves making the most of information contained in the order statistics obtained from the observations. Basically, the approach equates (matches) the model cdf evaluated at the observed order statistics to their corresponding percentile values. This system of equations can then be solved for the models distributional parameters.

In the case of the GPD model for the severity, there are two model parameters so one requires two distinct order statistics as a minimum to perform the estimation. Consider a set of realized data obtained under a POT's approach, where the n losses that have exceeded a pre-specified threshold level u are denoted by the data $\{x_i\}_{i=1:n}$ with order statistics denoted by $\{x_{(i)}(n)\}_{i=1:n}$. Given the cdf of the GPD model in Equation 2.192

$$F(x; \gamma, \sigma) = \begin{cases} 1 - (1 - (\gamma x/\sigma))^{1/\gamma}, \ \gamma \neq 0, \sigma > 0, \\ 1 - \exp(-x/\sigma), \ \gamma = 0, \sigma > 0, \end{cases} \tag{2.192}$$

we match the cdf at two of the randomly selected order statistics $i \neq j \in \{1, 2 \ldots, n\}$ to the corresponding percentile values,

$$\begin{aligned} F(x_{(i)}(n); \gamma, \sigma) &= p_{(i)}(n) \\ F(x_{(j)}(n); \gamma, \sigma) &= p_{(j)}(n), \end{aligned} \tag{2.193}$$

where the percentile is given for the GPD model with n observations by

$$p_{(i)}(n) = \frac{i - \eta}{n + \zeta}. \tag{2.194}$$

It is recommended in Castillo & Hadi (1997) that choices of $\eta = 0$ and $\zeta = 1$ provide reasonable results.

The solution to this system of equations in terms of the parameters is obtained by solving the equations for γ and σ given by

$$\begin{aligned} 1 - \left(1 - \frac{\widehat{\gamma} x_{(i)}(n)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{\gamma}}} &= \frac{i}{n+1}, \\ 1 - \left(1 - \frac{\widehat{\gamma} x_{(j)}(n)}{\widehat{\sigma}}\right)^{\frac{1}{\widehat{\gamma}}} &= \frac{j}{n+1}. \end{aligned} \tag{2.195}$$

Hence, for any two pairs of order statistics i, j, the solutions to these system of equations is given by

$$\begin{aligned} \widehat{\gamma}(i,j) &= \frac{\ln\left(1 - \left(x_{(i)}(n)/\widehat{\delta}(i,j)\right)\right)}{C_i} \\ \widehat{\sigma}(i,j) &= \widehat{\gamma}(i,j)\widehat{\delta}(i,j), \end{aligned} \tag{2.196}$$

in terms of $C_i = \ln(1 - p_{(i)}(n)) < 0$ and $\widehat{\delta}(i,j)$. Here $\widehat{\delta}(i,j)$ is the solution to the equation

$$C_i \ln\left(1 - \frac{x_{(j)}(n)}{\delta}\right) = C_j \ln\left(1 - \frac{x_{(i)}(n)}{\delta}\right), \tag{2.197}$$

which is obtained using a univariate root finding algorithm such as bisection. Note that δ corresponds to a reparameterizations of the GPD when $\delta = \sigma/\gamma$.

Remark 2.20 (EPM and Pickands Analytic Solution) *A special case of the EPM estimators is widely used in estimation of the GPD model parameters and known as the Pickands estimators. These correspond to the EPM setting in which $i = n/2$ and $j = 3n/4$. In these special cases, the bisection method is not required as the system of equations can be solved analytically according to the equation*

$$\begin{aligned} \widehat{\gamma} &= \tfrac{1}{\ln 2} \ln\left(\frac{x_{(n/2)}(n)}{x_{(3n/4)}(n) - x_{(n/2)}(n)}\right) \\ \widehat{\sigma} &= \widehat{\gamma}\left(\frac{x^2_{(n/2)}(n)}{2x_{(n/2)}(n) - x_{(3n/4)}(n)}\right). \end{aligned} \tag{2.198}$$

In general, one would not just pick two indexes i, j and instead would combine the Algorithms 1 and 2 discussed in Castillo & Hadi (1997) to produce an estimate of the GPD parameters. This involves the following sequence of steps.

Algorithm 2.11 (Elemental Percentile Estimation for GPD Parameter Estimation) *Combining Algorithm 1 and Algorithm 2 of Castillo & Hadi (1997).*

1. *Repeat for all order indexes $\{i, j : i < j, \text{for } i, j \in \{1, 2, \ldots, n\}\}$, such that $x_{(i)}(n) < x_{(j)}(n)$, the following steps:*
   (a) *Compute for $s \in \{i, j\}$ the values $C_s = \ln\left(1 - \frac{s-\eta}{n+\zeta}\right)$.*
   (b) *Set $d = C_j x_{(i)}(n) - C_i x_{(j)}(n)$, if $d = 0$, let $\widehat{\delta}(i,j) = \pm\infty$ and set the EVI estimate $\widehat{\gamma}(i,j) = 0$, otherwise compute $\delta_0 = x_{(i)}(n)x_{(j)}(n)(C_j - C_i)/d$.*

(c) If $\delta_0 > 0$, then $\delta_0 > x_{(j)}(n)$ and the bisection method can be used for the interval $[x_{(j)}(n), \delta_0]$ to obtain a solution $\widehat{\delta}(i,j)$. Otherwise the bisection method is applied to the interval $[\delta_0, 0]$.

(d) Use $\widehat{\delta}(i,j)$ to compute $\widehat{\gamma}(i,j)$ and $\widehat{\sigma}(i,j)$ using

$$\widehat{\gamma}(i,j) = \frac{\ln\left(1 - \dfrac{x_{(i)}(n)}{\widehat{\delta}(i,j)}\right)}{C_i} \tag{2.199}$$

$$\widehat{\sigma}(i,j) = \widehat{\gamma}(i,j)\widehat{\delta}(i,j).$$

2. *Take the median of each of the sets of estimated parameters for the overall estimator to obtain*

$$\begin{aligned} \widehat{\gamma}^{EPM} &= \mathbb{M}\text{edian}\{\widehat{\gamma}(1,2), \widehat{\gamma}(1,3), \ldots, \widehat{\gamma}(n-1,n)\} \\ \widehat{\sigma}^{EPM} &= \mathbb{M}\text{edian}\{\widehat{\sigma}(1,2), \widehat{\sigma}(1,3), \ldots, \widehat{\sigma}(n-1,n)\}. \end{aligned} \tag{2.200}$$

It is worth noting that, generally, in the OpRisk setting in which we consider heavy-tailed severity models, the amount of data obtained exceeding a given threshold u is relatively small. Therefore, in small sample sizes, n small, one may calculate all these combination of order statistics without difficulty. Otherwise, a subset of pairs of order statistics is obtained through random selection.

2.6.3.2 Threshold Exceedance Model Estimation under OpRisk Loss Data Departures from the GPD Model. In this section, we consider the modeling of threshold exceedances of OpRisk losses in settings in which the loss exceedances are not statistically well approximated as being distributed from a GPD exceedance model. It was discussed in Rootzén & Tajvidi (1997) that the standard estimators proposed in the literature for estimation of extremes (typically based on likelihood or MOM) are typically not robust to deviations of the loss data from the distributional assumptions made in developing the GPD extremes model.

In Feuerverger & Hall (1999), two semi-parametric methods are developed for modeling exceedance data, which will be suitable in the settings in OpRisk in which the loss exceedances depart from a GPD. To illustrate this, we reconsider the examples developed in Section 2.5.1.2 and consider losses which depart from the assumption discussed in that subsection relating to slowly varying tail properties. Often such departures from the theoretical properties of the GPD model are observed in practical loss modeling scenarios, so it is important to also consider methods robust to such departures.

In particular, in Feuerverger & Hall (1999), they present the situation in which the left tail is studied, although the approaches considered extend naturally to the right tail we are typically studying in this chapter. It is assumed that the model from which the data was drawn has a distribution $F(x)$, which satisfies

$$F(x) = Cx^{\alpha}\{1 + \delta(x)\}, \tag{2.201}$$

where $C, \alpha > 0$ and $\lim_{x \downarrow 0} \delta(x) \to 0$ and the focus on estimation is on the tail index α. Then as in Section 2.5.1.2, we would like to assume that the tails are slowly varying and, therefore, that the GPD model is suitable; therefore, the distribution is approximately

$$F(x) \approx Cx^{\alpha}. \tag{2.202}$$

However, in cases in which this is not valid as we do not want to work with small sample sizes (very deep in the tail where GPD may be reasonable) or just because of deviations from the assumptions required for the POT's methodology, then we require robust estimation methods for the GPD model in such scenarios.

The two proposed methods developed in Feuerverger & Hall (1999) are based on approximate likelihood methods and least squares estimation. In particular, attention is paid to the estimation of the tail index (EVI) under departures of the loss data from the GPD EVT model assumptions discussed earlier. However, one should note that such approaches that will be developed later should only be utilized in cases in which the loss data is believed to deviate from the standard POT's assumptions for the GPD model. This is due to the fact that the developed techniques result in a significant variance increase in parameter estimations under scenarios in which conventional extreme value approximations are appropriate. However, when modeling exceedances in the presence of departures of the assumptions discussed regarding GPD properties, the developed techniques are able to reduce bias by an order of magnitude without a variance inflation.

In particular, the model in Equation 2.201 can be interpreted as a perturbed GPD model presented in Equation 2.202. One can then consider a departure from the GPD model obtained by modeling the function $\delta(x)$ via, for example,

$$\delta(x) = Dx^{\beta} + o(x^{\beta}), \text{as } x \to 0, \tag{2.203}$$

with $\beta > 0$ and $-\infty < D < \infty$. Such a model turns out to be equivalent to consideration of a model for the loss severities, which is a mixture of two Pareto distributions up to first order for the departures, given by

$$F(x) = C_1 x^{\alpha_1} + C_2 x^{\alpha_2}, \tag{2.204}$$

with $x \in [0, \varepsilon)$, C_1, α_1, α_2 are positive and identifiable when $\alpha_1 < \alpha_2$. Other departures can also be considered under the proposed methodology to be presented, and before developing this estimation, it is important to mention a simple diagnostic tool to assess the presence of such departures.

Remark 2.21 (Assessing Departures in Exceedance Losses from the GPD Model) *In practice, there are simple visual checks that risk modellers can perform on the loss data to assess quickly the possibility of departures from GPD models and, therefore, when to utilize the following estimation procedures. The first of these is to consider the goodness-of-fit of the GPD model to the tail of the distribution of the losses through a plot of the log transform of the extreme order statistics against the log of their ranks. If the linear relationship is plausible, the negative of the gradient is a simple estimator of the EVI.*

If one considers modeling data from the tail of the empirical distribution of the observed losses, and there is marked deviation from the GPD model assumptions, in this case, the results from the estimation of the GPD model tail exponent can be misleading. In addition, if one goes far enough into the tails of the distribution, fitting using the small samples from the very high order statistics, then the GPD model may be more reasonable; however, the variance in the estimates will be relatively very large. To reduce the bias that arises if one models the extreme value data relatively deeply into the tails and, in addition, there is deviation from the GPD model assumptions and, therefore, distributional assumptions, one can consider the proposal of Feuerverger & Hall (1999) to consider log spacings of the order statistics. This was considered in two approaches: the first based on a likelihood approximation and the second based on a

least squares regression. The bias reduction approach for the regression is based on the log–log spacings of order statistics in the tail (exceedances) being treated as response variables and the covariates in the linear model are the ranks of the order statistics; to complete the model, the regression errors are assumed to have a type III extreme value distribution with exponentially light tails.

2.6.3.3 Log-Spaced Order Statistic Approximate Likelihood Robust Estimation of the GPD Tail Index. In the following, we present the likelihood-based estimation for the EVI for loss data that deviates from GPD model assumptions.

Algorithm 2.12 (Robust EVI Estimation) *Under Approximate Likelihood of Normalized Log-Spaced small (High) Order Statistics.*

1. *Sort the n observed losses $\{X_i\}_{i=1:n}$ to obtain order statistics $\{X_{(i)}\}_{i=1:n}$.*
2. *Evaluate the log differences to obtain the series $\{U_i : U_i = i(\ln X_{(i+1)} - \ln X_{(i)})\}_{i=1:n}$.*
3. *Consider a function $\delta_1(x)$ which can be expressed in terms of the function in Equation 2.201 and consider*

$$U_i \approx Z_i \frac{1}{\alpha}\{1 + \delta_1(i/n)\} \approx Z_i \frac{1}{\alpha}\exp(\delta_1(i/n)) \tag{2.205}$$

   with i.i.d. $Z_i \sim \text{Exp}(1)$.
4. *If $\delta(x)$ in Equation 2.201 satisfies model specification in Equation 2.203, then one can present*

$$\delta_1(x) = D_1 x^{\beta_1} + o(x^{\beta_1}), \text{as } x \downarrow 0 \tag{2.206}$$

   and $\beta_1 = \beta/\alpha$ and $D_1 = -(\beta/\alpha)\, C^{-\beta/\alpha} D$.
5. *Hence, if one approximates the U_i random variables as $U_i \sim \text{Exp}(D_1(i/n)^{\beta_1})$ and the parameters α, D_1, β_1 can be estimated by maximizing the resulting approximate likelihood given by a negative log-likelihood in Equation 2.207,*

$$\ln L(\mathbf{U}; \alpha, D_1, \beta_1) = -r\ln\alpha + D_1 \sum_{i=1}^{r} (i/n)^{\beta_1} + \alpha \sum_{i=1}^{r} U_i \exp(-D_1(i/n)^{\beta_1}), \tag{2.207}$$

   where r is the smoothing parameter.

2.6.3.4 Log–Log Spaced Order Statistics Regression-Based Robust Estimation of the GPD Tail Index. Alternatively, the least-squares-based regression approach simply involves minimization of the sum of squared error criterion

$$SS = \sum_{i=1}^{r} (V_i + \ln\alpha - \mu_0 - D_1(i/n)^{\beta_1})^2, \tag{2.208}$$

with log–log spacings $V_i = \ln U_i \approx -\ln\alpha + \mu_0 + \delta_1(i/n) + \varepsilon_i$, μ_0 is the mean of the distribution of $\ln Z_1$ (the negative of Euler's constant) and $\varepsilon_i = \ln Z_i - \mu_0$ (for $1 \le i \le n$), which are the approximation errors in the regression model.

2.6.3.5 Robust Estimation of the GPD Tail Index via Method of Medians. In Peng & Welsh (2001), an alternative robust estimation procedure for the GPD model is also proposed based on the method of medians approach. This was originally developed for Weibull distributions in He & Fung (1999) and is extended to estimate the parameters of the GPD for threshold exceedances. It involves equating the sample median of each component of the score function to the population median.

Definition 2.32 *The score function in the GPD model with a density given by*

$$f_X(x;\gamma,\sigma) = \frac{1}{\sigma}\Big(1+\frac{\gamma x}{\sigma}\Big)^{-\frac{1}{\gamma}-1}\left\{\mathbb{I}\left[x>0,\gamma\geq 0\right]+\mathbb{I}\left[0<x<-\frac{\sigma}{\gamma},\gamma<0\right]\right\} \tag{2.209}$$

involves the following two derivatives of the log density,

$$\begin{aligned}\frac{\partial}{\partial\sigma}\ln f_X(x;\gamma,\sigma) &= \left(-\frac{1}{\sigma}+\frac{1+\gamma}{\sigma^2}\frac{x}{1+\frac{\gamma x}{\sigma}}\right)\\ &\quad\times\left\{\mathbb{I}\left[x>0,\gamma\geq 0\right]+\mathbb{I}\left[0<x<-\frac{\sigma}{\gamma},\gamma<0\right]\right\}\\ \frac{\partial}{\partial\gamma}\ln f_X(x;\gamma,\sigma) &= \left(\frac{\ln\Big(1+\frac{\gamma x}{\sigma}\Big)}{\gamma^2}-\frac{(1+\gamma)x}{\sigma\gamma+\gamma^2 x}\right)\\ &\quad\times\left\{\mathbb{I}\left[x>0,\gamma\geq 0\right]+\mathbb{I}\left[0<x<-\frac{\sigma}{\gamma},\gamma<0\right]\right\}.\end{aligned} \tag{2.210}$$

Given these score functions for the GPD, one can apply the method of medians for the first component of the score function to obtain

$$\sigma = \frac{\gamma}{2^\gamma-1}\mathbb{M}\text{edian}\,(X_i), \tag{2.211}$$

where $\mathbb{M}\text{edian}(X_i)$ is the sample median of $\{X_i\}_{i=1:n}$. The method of medians for the second component of the score function provides the second equation

$$\mathbb{M}\text{edian}\left(\frac{\ln(1+(\gamma X_i/\sigma))}{\gamma^2}-\frac{(1+\gamma)X_i}{\sigma\gamma+\gamma^2 X_i}\right) = z(\gamma), \tag{2.212}$$

with $z(\gamma)$ satisfying the equation

$$\int_{\left\{0<y<1,-\frac{\ln y}{\gamma}-\frac{1+\gamma}{\gamma^2}(1-y^\gamma)>z(\gamma)\right\}} dy = \frac{1}{2}. \tag{2.213}$$

Hence, the estimators of the EVI and scale parameters of the GPD are obtained through the solutions to Equations 2.211 and 2.212.

2.6.4 EVT—RANDOM NUMBER OF LOSSES

As has already been mentioned, in OpRisk, the number of losses per time period is not fixed and is modeled as a random variable N with $p_n = \Pr[N = n]$. This has some implications for the use of the above-described EVT.

Consider a sequence of n independent random variables $X_1, \ldots, X_n$ from a distribution $F(x)$ representing losses, where n is a known number. Denote the maximum loss as

$$M_n = \max(X_1, \ldots, X_n).$$

Because each loss cannot exceed the maximum and owing to independence between the losses, the distribution of the maximum is

$$\begin{aligned} F_{M_n}(x) &= \Pr[M_n \leq x] = \Pr[X_1 \leq x, \ldots, X_n \leq x] \\ &= \prod_{i=1}^{n} \Pr[X_i \leq x] = (F(x))^n. \end{aligned} \tag{2.214}$$

If the frequency is random, consider a sequence of N independent random variables $X_1, \ldots, X_N$ from a distribution $F(x)$, where N is random. Denote the maximum loss as

$$M_N = \max(X_1, \ldots, X_N),$$

then, instead of Equation 2.214, the distribution of a maximum M_N is calculated as

$$\begin{aligned} F_{M_N}(x) &= \sum_{n=0}^{\infty} \Pr[M_N \leq x | N = n] \Pr[N = n] \\ &= \sum_{n=0}^{\infty} (F(x))^n \Pr[N = n] = \psi_N(F(x)), \end{aligned} \tag{2.215}$$

where

$$\psi_N(t) = \mathrm{E}[t^N] = k\sum p_k t^k$$

is the probability generating function of the frequency distribution. Note that there is a finite probability for zero maximum, that is, $\Pr[M_N = 0] = \psi_N(F(0))$. Typically, severity distribution has $F(0) = 0$ and frequency distribution has a finite probability at zero; in this case, $\Pr[M_N = 0] = \Pr[N = 0]$.

EXAMPLE 2.6 (Distribution of Maximum for Poisson Frequency)

If the annual number of losses $N \sim Poisson(\lambda)$, then the probability generating function of the frequency is

$$\psi(t) = \exp(-\lambda(1 - t))$$

and thus the distribution of the maximum loss (per annum) is

$$F_{M_N}(x) = \exp(-\lambda(1 - F(x)). \tag{2.216}$$

The distribution of the maximum loss over m years is

$$(F_{M_N}(x))^m = \exp(-m\lambda(1 - F(x)). \tag{2.217}$$

■

EXAMPLE 2.7 (Distribution of Maximum for Poisson Frequency and GPD Severity)

If the annual number of losses $N \sim Poisson(\lambda)$ and the severities are from $GPD(\xi, \beta)$, then in the case of $\xi \neq 0$,

$$\begin{aligned} F_{M_N}(x) &= \exp(-\lambda(1 + \xi x/\beta)^{-1/\xi}) \\ &= \exp(-(1 + \xi(x - \mu)/\sigma)^{-1/\xi}), \quad x \geq 0, \end{aligned} \tag{2.218}$$

where $\sigma = \beta\lambda^{\xi}$ and $\mu = (\beta\lambda^{\xi} - \beta)/\xi$. In the case of $\xi = 0$,

$$\begin{aligned} F_{M_N}(x) &= \exp(-\lambda \exp(-x/\beta)) \\ &= \exp(-\exp(-(x - \mu)/\sigma)), \quad x \geq 0, \end{aligned}$$

where $\mu = \beta \ln \lambda$ and $\sigma = \beta$. Thus the distribution of M_N is a three-parameter GEV $H_{\xi,\mu,\sigma}(x)$ for $x \geq 0$ and zero for $x < 0$. Note that $F_{M_N}(x)$ is continuous at $\xi = 0$. It is important to note that $F_{M_N}(x)$ is not GEV for all x but for $x \geq 0$ and there is a finite probability at zero $\Pr[M_N = 0] = \exp(-\lambda)$. If λ increases, then $\Pr[M_N = 0] \to 0$ and $F_{M_N}(x)$ will converge to GEV on the whole domain. In the case of other severity distributions, one can consider the limit of large λ when the distribution of $F_{M_N}(x)$ converges to a continuous distribution function, then it can be argued that the limiting distribution of maxima over many time periods is GEV. ■

Frequency of exceedances. Consider N independent losses $X_1, \ldots, X_N$ with a common distribution $F(x)$, where N is a discrete random variable with $p_n = \Pr[N = n]$. Then the number of losses above a threshold u is

$$N_u = I_1 + \cdots + I_N,$$

where I_j are independent indicator random variables. Then, the probability-generating function of N_u can be calculated using probability-generating function of N, $\psi_N(t)$, and probability- generating function of I_j, $\psi_I(t)$, as

$$\psi_{N_u} = \psi_N(\psi(I)).$$

Moreover, in the case when N is from Poisson, binomial or negative binomial, the distribution of N_u is the same as distribution of N with only one parameter changed.

- *Poisson.* If the frequency of losses is from $Poisson(\lambda)$, then the frequency of losses above u is $Poisson(\widetilde{\lambda})$ with $\widetilde{\lambda} = \lambda(1 - F(u))$.

- *Negative Binomial.* If the frequency of losses is from $NegBin(r, p)$, where the parameter $p = 1/(1+q)$, then the frequency of losses above u is $NegBin(r, \widetilde{p})$ with $\widetilde{p} = 1/(1+\widetilde{q})$, where $\widetilde{q} = q(1 - F(u))$.
- *Binomial.* If the frequency of losses is from $Bin(n, p)$, then the frequency of losses above L is $Bin(n, \widetilde{p})$, where $\widetilde{p} = p(1 - F(u))$.

It can be argued that binomial and negative binomial distributions will converge to Poisson when u increases, that is, when $1 - F(u) \to 0$. This is intuitively expected because in this limit, the variance and mean of these distributions converge to the mean of Poisson. There is an extensive literature on point processes of extremes, see McNeil *et al.* (2005, section 7.4) or Embrechts *et al.* (1997, chapter 5). For the purposes of this book, we just say that, in general, it is argued that the distribution of the exceedances over a high threshold is well approximated by Poisson distribution. Moreover, the theory suggests to model the high threshold exceedances by the Poisson process.

Remark 2.22 *The EVT threshold exceedances says that for a large threshold, the distribution of loss exceedances $F(x)$ can be approximated by GPD. Thus one can always argue that the distribution of the maximum of i.i.d. excesses over a high threshold can be approximated by GEV in the case of Poisson frequency, see Equation 2.218. The assumption of Poisson distribution for the frequency of exceedances is also typical for high threshold exceedances.*

Chapter Three

Heavy-Tailed Model Class Characterizations for LDA

In this chapter, several important concepts that will be utilized throughout the following chapters will be developed and presented. The chapter will start with an introduction to the class of asymptotic notations that are adopted widely in the actuarial and mathematical literature. Then a high level presentation of the key components of characterizing tails of distributions will be presented. In particular, we will consider the many varied ways that one may classify a distribution function as being heavy tailed, this will allow a solid mathematical footing to be placed on the discussion about LDA models with heavy-tailed severity model and the effect this may have on modeling and capital estimation. We do not aim to provide a comprehensive coverage of this topic as there are already many great books on this topic, see, for instance, Bingham *et al.* (1989) and the more recent book by Borovkov and Borovkov (2002) to name a few. Instead we will provide in this chapter a guided tour through some important key results that will be applicable to both practitioners and researchers alike when investigating key aspects of capital estimation for operational risk (OpRisk) settings through asymptotic analysis. The following key topics are discussed:

1. Asymptotic notations;
2. Heavy-tailed models in the sub-exponential distribution class;
3. Heavy-tailed models in the regular and slow variation distribution classes;
4. Refinements of heavy-tailed model classes:
   - Consistent variation;
   - Smooth variation;
   - Dominant variation;
   - Subversive variation;
   - Asymptotically locally constant tails (long tailed);
5. Extended regular variation and Matuszewska indices.

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

3.1 Landau Notations for OpRisk Asymptotics: Big and Little 'Oh'

In this section, the closed-form approximations are provided for single-loss processes in which dependence is not present. To begin this section, one needs to be familiar with some important notations that will be used routinely throughout the different representations developed. The Landau notation is utilized to represent asymptotic growth of functions and can be denoted by either little-Oh $o(\cdot)$ or big-Oh $O(\cdot)$, as defined in the following definitions. In Definition 3.1, we present the concept of the asymptotic upper bound of a function; this is followed by the definition of the asymptotic lower bound of a function in Definition 3.2.

Definition 3.1 (Big-Oh Asymptotic Upper Bound) *Let $f(x)$ and $g(x)$ be two functions defined on a subset of $\mathbb{R}$. Then the set $O(g(x))$ is defined as*

$$O(g(x)) = \{f(x) | \exists c > 0, \exists x_0 > 0, \forall x \geq x_0 : 0 \leq f(x) \leq cg(x)\}, \tag{3.1}$$

where $g(x)$ is an asymptotic upper bound for $f(x)$.

In other words, $f(x) \in O(g(x))$ if and only if there exist positive constants c, and x_0, such that for all $x \geq x_0$, the inequality $0 \leq f(x) \leq cg(x)$ is satisfied.

Definition 3.2 (Asymptotic Lower Bound) *Let $f(x)$ and $g(x)$ be two functions defined on a subset of $\mathbb{R}$. Then the set $\Omega(g(x))$ is defined as*

$$\Omega(g(x)) = \{f(x) | \exists c > 0, \exists x_0 > 0, \forall x \geq x_0 : 0 \leq cg(x) \leq f(x)\}, \tag{3.2}$$

where $g(x)$ is an asymptotic lower bound for $f(x)$.

In other words, $f(x) \in \Omega(g(x))$ if and only if there exist positive constants c, and x_0, such that for all $x \geq x_0$, the inequality $0 \leq cg(x) \leq f(x)$ is satisfied.

Corollary 3.1 *A consequence of this is that $f(x) = O(g(x))$ if and only if $g(x) = \Omega(f(x))$.*

We may also define the less informative asymptotic notation given by little-Oh $o(\cdot)$ notation as detailed in Definition 3.3. This definition looks similar to that of the big-Oh asymptotic upper bound, but the key difference is in the definition of the constant c; whilst the big-Oh result states the existence of a value c that multiplies a function to obtain an upper bound, the little-Oh notation states that this will hold for all c and is therefore less informative on rates of convergence.

Definition 3.3 (Little-Oh Strict Asymptotic Upper Bound) *Let $f(x)$ and $g(x)$ be two functions defined on a subset of $\mathbb{R}$. Then the set $o(g(x))$ is defined as*

$$o(g(x)) = \{f(x) | \forall c > 0, \exists x_0 > 0, \forall x \geq x_0 : 0 \leq f(x) \leq cg(x)\}, \tag{3.3}$$

where $g(x)$ is a strict asymptotic upper bound for $f(x)$.

As a consequence of this definition, we may write the Lemma 3.1 relating to the limiting behaviour of the ratio of the two functions.

Lemma 3.1 *The statement $f(x) = o(g(x))$ is true if and only if*

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0. \tag{3.4}$$

Furthermore, it will be useful to also note the following properties of the Landau big-Oh notation given in Lemma 3.2, see detailed discussion in De Bruijn (1970).

Lemma 3.2 (Properties of Big-Oh Notations) *The following properties of big-Oh notation apply (where explicitly stated they also apply to little-Oh notation):*

1. ***Multiplication by Constants.*** *Given a positive constant $(c > 0)$ if $f(x) \in O(g(x)) \Rightarrow cf(x) \in O(g(x))$, therefore, one has the equivalence between the statements $f(x) = O(c)$ and $f(x) = O(1)$. This also applies for little-Oh notation.*
2. ***Transitivity.*** *Given $f(x) = O(g(x))$ and $g(x) = O(h(x))$, then one has $f(x) = O(h(x))$. This also applies for little-Oh notation.*
3. ***Multiplication of Functions.*** *Given two functions with asymptotic bounds according to $f_i(x) = O(g_i(x))$ for $i \in \{1, 2\}$ (equivalently $f_1(x) \in O(g_1(x))$) and $f_2(x) = O(g_2(x))$, then the product of these two functions satisfies asymptotically the bound given by $f_1(x)f_2(x) = O(g_1(x)g_2(x))$. This also applies for little-Oh notation.*
4. ***Factorization of Functions.*** *Given $f(x) = O(g(x)h(x))$ then one can factorize according to $f(x) = g(x)O(h(x))$ to remove important terms from big-Oh expressions. This also applies for little-Oh notation.*
5. ***Summation of Functions.*** *Given $f_i(x) = O(g_i(x))$, then the linear combination satisfies*

$$\sum_{i=1}^{n} f_i(x) = O\left(\sum_{i=1}^{n} |g_i(x)|\right). \tag{3.5}$$

6. ***Integration of Functions.*** *Given to integrable functions on finite intervals $f(x)$ and $g(x)$ with $f(x) = O(g(x))$ for $x \geq x_0$, then one has*

$$\int_{x_0}^{x} f(\tau)d\tau = O\left(\int_{x_0}^{x} |g(\tau)|d\tau\right), \ \forall x \geq x_0. \tag{3.6}$$

Remark 3.1 *It is important to note that when interpreting equations which comprise "Oh" notations on either side of an equality, are not symmetric, instead they should be interpreted from left to right. For example, take the relationship for the domain $x \geq 1$ given by $O(\sqrt{x}) = O(x)$, then one interprets this to mean that if $f(x) = O(\sqrt{x})$ then of course it also satisfies $f(x) = O(x)$, clearly the reverse is not true in general.*

Remark 3.2 *One should also observe that O-terms in equations on either side of an equality* ***cannot be cancelled****. For example, given $f(x) = h(x) + O(d(x))$ and $g(x) = h(x) + O(d(x))$, one can only conclude that $f(x) = g(x) + O(d(x))$ in which the two functions $f(x)$ and $g(x)$ differ at most asymptotically by a function of order $O(d(x))$.*

Finally, we remark on the use of $o(\cdot)$ and $O(\cdot)$ where we note that a little-Oh estimate provides a qualitative statement, where $f(x) = o(g(x))$ tells one that the quotient $f(x)/g(x)$

tends to 0 as $x \to \infty$. However, this statement is not quantitative as it does not provide information on the rate of this convergence, which can often be important to understand. Therefore, when possible one may alternatively provide asymptotic relations with respect to a big-Oh estimate. Using the big-Oh estimate for the order of asymptotic approximation implies that one has a more precise functional characterization of the behaviour $f(x) = O(g(x)/h(x))$ with some explicit function $h(x)$ that tends to infinity as $x \to \infty$ and, consequently, implies the litte-Oh estimate $f(x) = o(g(x))$, clearly therefore providing more information.

In the following, several practically useful examples of simplification using big-Oh notation are provided.

EXAMPLE 3.1

One can show using the results in Lemma 3.2 that for $g(x) \to 0$ as $x \to \infty$, the following simplifications hold:

$$\begin{aligned}
\exp(O(\,g(x))) &= 1 + O(\,g(x)) \\
\ln(1 + O(\,g(x))) &= O(\,g(x)) \\
(1 + O(\,g(x)))^p &= 1 + O_p(\,g(x)), \text{ for any } p \in \mathbb{R}, \\
(1 + O(\,g(x)))^{-1} &= 1 + O(\,g(x)).
\end{aligned}$$

Also one can show the relationship between big-Oh and little-Oh for a Taylor series example as follows.
Let $f : \mathbb{R} \mapsto \mathbb{R}$ be an $(n+1)$-times continuously differentiable function. Then the Taylor series expansion of order n is given by

$$f(x) = \left[\sum_{k=0}^{n} \frac{1}{k!} \frac{d^k}{dx^k} f(x) \bigg|_{x=x_0} (x - x_0)^k \right] + O((x - x_0)^{n+1}).$$

Furthermore, if the function $f(x)$ is only n-times continuously differentiable, instead of $(n+1)$-times, then one has that

$$f(x) = \left[\sum_{k=0}^{n} \frac{1}{k!} \frac{d^k}{dx^k} f(x) \bigg|_{x=x_0} (x - x_0)^k \right] + o((x - x_0)^n).$$

■

We conclude this brief introduction to asymptotic notations by commenting on alternative closely related notations that have been adopted in the risk and insurance literature, primarily discussing the Vinagradov notations given in Definition 3.4.

Definition 3.4 (Vinogradov Notations) *Consider $x \in \mathbb{R}$ and consider the interval of values $x \geq x_0$ for some x_0, then the following equivalences between Landau and Vinagradov notations applies for two functions $f(x)$ and $g(x)$:*

1. *The statement that $f(x)$ and $g(x)$ have the same order of magnitude, denoted by $f(x) \approx g(x)$, is interpreted to mean that $f(x) = O(\,g(x))$ and $g(x) = O(\,f(x))$. This statement on the order of magnitude of two functions holds iff $\exists c_1, c_2 > 0$ and $\exists x_0$ s.t.*

$$c_1|g(x)| \leq |f(x)| \leq c_2|g(x)|, \quad x \geq x_0. \tag{3.7}$$

2. *The statement* $f(x) \ll g(x)$ *is interpreted to mean that* $f(x) = O(g(x))$.

3. *The statement* $f(x) \gg g(x)$ *is equivalent to the statement* $g(x) \ll f(x)$.

Next we return to the Landau "Oh" notations and provide a few simple examples to illustrate how to manipulate the Landau notations.

EXAMPLE 3.2

To illustrate the difference between big-Oh and little-Oh notations, consider a sequence of functions $\{f(n)\}$ such that

$$|f(n) - 4| \leq \frac{8}{n}, \quad n = 1, 2, 3, \ldots \tag{3.8}$$

Then it is clear that as $n \to \infty$ one must have that

$$f(n) \to 4. \tag{3.9}$$

However, in many settings, the knowledge that $f(n)$ eventually becomes 4 is not sufficient, one instead would like to know the rate, that is, how quickly does $f(n)$ converge to 4. In addition, it may often be the case that the knowledge that $f(n)$ will converge to 4 with the multiplier 8 in the bound would be just as useful as any other constant being there so it could be suitable to simply state that $\exists C \in \mathbb{R}$ independent of n s.t.

$$|f(n) - 4| \leq \frac{C}{n}, \quad n = 1, 2, 3, \ldots \tag{3.10}$$

This last statement can be written equivalently in big-Oh notation according to

$$f(n) - 4 = O\left(\frac{1}{n}\right), \quad n = 1, 2, 3, \ldots \tag{3.11}$$

Then one can see that Equation 3.9 would imply Equation 3.11, which in turn would imply Equation 3.10. ■

Remark 3.3 *So one can see that the big-Oh notation can be read in words as meaning:*

'Something that is in absolute value less than a constant number of times'

The little-Oh notation should be read as meaning:

'Something that tends to (...constant...), multiplied by...'

Note that little-Oh's $o(\cdot)$ *are less important than big-Oh's* $O(\cdot)$ *since they carry less information, they do not provide the rate at which convergence is achieved, only that it can be achieved.*

Before, proceeding with sections which define different classes of heavy-tailed models by the variation of their tails, it will also be important to introduce two additional basic concepts for OpRisk LDA tail asymptotic analysis, namely, asymptotic equivalence and max-sum equivalence.

Definition 3.5 (Asymptotic Equivalence) *A probability distribution function $F(x)$ is asymptotically equivalent to another probability distribution function $G(x)$, denoted by $F(x) \sim G(x)$ as $x \to \infty$ if it holds that $\forall \varepsilon > 0, \exists x_0$ such that $\forall x > x_0$, the following is true:*

$$\left| \frac{F(x)}{G(x)} - 1 \right| < \varepsilon. \tag{3.12}$$

To define the notion of max-sum equivalence, it will be important to first recall the definition of the tail convolution, see Definition 3.6.

Definition 3.6 (Convolution of Distributions and Tail Convolutions) *For any given Borel set B, one can define the distribution of the sum of two independent loss random variables X_1 and X_2, with distributions F and G, respectively, according to convolution between the two distributions F and G given by*

$$(F \star G)(B) = \mathbb{P}\mathrm{r}[X_1 + X_2 \in B] = \int_{-\infty}^{\infty} F(B-y)\,dG(y) = \int_{-\infty}^{\infty} G(B-y)\,dF(y). \tag{3.13}$$

where $F(B-y)$ is interpreted as the set of values of $x \in B$ and the argument is the set minus elements of y. Furthermore, the convolution tail of F and G is then given by

$$\overline{F \star G}(x) = \mathbb{P}\mathrm{r}[X_1 + X_2 > x] = \int_{-\infty}^{\infty} \overline{F}(x-y)\,dG(y) = \int_{-\infty}^{\infty} \overline{G}(x-y)\,dF(y). \tag{3.14}$$

We will use throughout the following definitions for lim-sup, lim-inf and oscillation of a function or sequence.

Definition 3.7 (Limit Supremum, Limit Infimum and Oscillation of a Function) *If a function $f(x)$ is defined with respect to metric spaces X and Y, where Y is an ordered set. Then for a subspace E contained in X if one has the function $f(x) : E \mapsto Y$, then one may define the $limsup_{x \to a}$ as follows*

$$\mathrm{limsup}_{x \to a} f(x) = \lim_{\varepsilon \to 0} (\sup\{f(x) : x \in E \cap B(a;\varepsilon) \backslash \{a\}\}),$$

where $B(a;\varepsilon)$ denotes the metric ball of radius ε around a. In addition, one may define the $liminf_{x \to a}$ as follows

$$\mathrm{liminf}_{x \to a} f(x) = \lim_{\varepsilon \to 0} (\inf\{f(x) : x \in E \cap B(a;\varepsilon) \backslash \{a\}\}).$$

The oscillation of a function $f(x)$ denoted by $osc(f(x))$ is given by the difference between the lim-sup and the lim-inf of the function at a point or in a set or interval. In general the oscillation of a function $f : E \subset \mathbb{R} \mapsto \mathbb{R}$ is given by

$$osc(f, E) = \sup_{a,b \in E} |f(b) - f(a)| \tag{3.15}$$

where a and b are arbitrary points in E. If one then defines a sequence of points or functions f_n one can define the $osc(f_n)$ as follows,

$$osc(f_n) = \text{limsup} f_n - \text{liminf} f_n. \tag{3.16}$$

Remark 3.4 (Limit Supremum) *The property that $limsup_{x\to\infty} f(x) = \infty$ can be interpreted as meaning that the values of $f(x)$ are unbounded from above on all neighbourhoods of $+\infty$, i.e. all open intervals of the form (a, ∞) for some $a \in \mathbb{R}$.*

Remark 3.5 (Oscillation of a Function) *The difference between the lim-sup and the lim-inf of a function $f(x)$ is an indication of how wildly the function $f(x)$ varies or oscillates.*

For real valued functions $f(x) : \mathbb{R} \mapsto \mathbb{R}$, then the limit inferior and limit superior are always well defined if we allow the values $+\infty$ and $-\infty$. For example consider the function $f(x) = \sin(1/x)$, then the $limsup_{x\to 0} f(x) = 1$ and the $liminf_{x\to 0} = -1$.

We can now start to think of several independent loss random variables being added together. In many cases we may not be able to find the solution to the n-fold convolutions for the distributions presented in Definition 3.6, therefore it will be often useful to think of some basic behaviours such as bounds. Therefore, it is interesting to observe that when one has n loss random variables which are independent but not necessarily identically distributed, the following bound can be shown to hold in Theorem 3.1.

Theorem 3.1 (Bounds on Aggregate Tail Distributions) *Consider n independent annual loss random variables $Z_1, Z_2, \ldots, Z_n$ each distributed according to $Z_i \sim F_i$ where each distribution is on a strictly positive support. One can bound the extreme tail probabilities of the aggregated total loss for*

$$Z_T = \sum_{i=1}^{n} Z_i, \tag{3.17}$$

by the sum of the individual annual loss tails according to

$$\text{liminf}_{z\to\infty} \frac{\overline{F_1 \star F_2 \star F_3 \star \cdots \star F_n}(z)}{\overline{F}_1(z) + \cdots + \overline{F}_n(z)} \leq 1. \tag{3.18}$$

To show this result, consider taking the following sets of disjoint events $\{Z_i > z, Z_j \in [0, z], \forall j \neq i\}$, then as discussed in Foss *et al.* (2011), one can then utilize these disjoint events to construct the following bounds on the aggregate tail distribution asymptotically as $z \to \infty$ according to

$$\begin{aligned}
\overline{F_1 \star \cdots \star F_n}(z) &\geq \sum_{i=1}^{n} \mathbb{P}\text{r}[Z_i > z, Z_j \in [0, z], \forall i \neq j] \\
&= \sum_{i=1}^{n} \overline{F}_i(z) \prod_{j\neq i} F_j(z) \\
&\sim \sum_{i=1}^{n} \overline{F}_i(z).
\end{aligned} \tag{3.19}$$

Given the definition of a tail convolution, we now define a special asymptotic result that will hold for certain distributions, known as *max-sum equivalence*. This is basically the above-stated bound with $n = 2$ in the case that equality holds; in later sections, we identify classes of heavy-tailed severity models for which this condition will be satisfied.

Definition 3.8 (Max-Sum Equivalence) *A probability distribution function is max-sum equivalent, denoted by $F \sim_M G$, when the convolution of the tail distribution of two random variables is distributed according to the sum of the two-tail distributions asymptotically,*

$$1 - (F \star G)(x) = (\overline{F \star G})(x) \sim \overline{F}(x) + \overline{G}(x), \quad x \to \infty,$$

see discussion in Li and Tang (2010).

To complete this section recalling some basic results, we also note the following two properties one may obtain as decompositions of the tail convolution identity:

- **Decomposition 1.** Consider the tail probability of the sum of two independent loss random variables $X_1 \sim F$ and $X_2 \sim G$ given by $\Pr[X_1 + X_2 > x]$. The following decomposition holds; for $\lambda > 0$ and any $x > 0$, one has

$$\overline{F \star G}(x) = \Pr[X_1 + X_2 > x, X_1 \leq \lambda] + \Pr[X_1 + X_2 > x, X_1 > \lambda], \tag{3.20}$$

  where one observes for instance that

$$\Pr[X_1 + X_2 > x, X_1 \leq \lambda] = \int_{-\infty}^{\lambda} \overline{G}(x - y)\, dF(y). \tag{3.21}$$

$$\begin{aligned} \Pr[X_1 + X_2 > x, X_1 > \lambda] &= \Pr[X_1 + X_2 > x | X_1 > \lambda] \Pr[X_1 + > \lambda] \\ &= \overline{F}(\lambda) \int_{\lambda}^{\infty} \overline{G}(x - y)\, dF(y). \end{aligned} \tag{3.22}$$

- **Decomposition 2.** Consider the tail probability of the sum of two independent loss random variables $X_1 \sim F$ an $X_2 \sim G$ given by $\Pr[X_1 + X_2 > x]$. The following decomposition holds, for any $x > 0$ with $\lambda \in [0, x/2]$, such that one obtains

$$\begin{aligned} \overline{F \star G}(x) = \Pr[X_1 + X_2 > x, X_1 \leq \lambda] + \Pr[X_1 + X_2 > x, X_2 \leq \lambda] \\ + \Pr[X_1 + X_2 > x, X_1 > \lambda, X_2 > \lambda] \end{aligned} \tag{3.23}$$

  where one observes that the probability of the event $\{X_1 + X_2 > x, X_1 > \lambda, X_2 > \lambda\}$ is symmetric thereby giving

$$\begin{aligned} \Pr[X_1 + X_2 > x, X_1 > \lambda, X_2 > \lambda] &= \int_{\lambda}^{\infty} \overline{F}(\max\{\lambda, x - y\})\, dG(y) \\ &= \int_{\lambda}^{\infty} \overline{G}(\max\{\lambda, x - y\})\, dF(y). \end{aligned} \tag{3.24}$$

These decompositions are particularly useful when discussing tail asymptotics of partial sums and resulting risk measure asymptotics, which we will see in future chapters.

3.2 Introduction to the Sub-Exponential Family of Heavy-Tailed Models

In probability theory, heavy-tailed distributions are probability distributions whose tails are not exponentially bounded: that is, they have heavier tails than the exponential distribution. In many risk and insurance contexts, the notion of a heavy-tailed distribution is obtained by considering those distributions F for which the right tail is sufficiently heavy that the following inequality in Equation 3.24 holds for all positive exponential moments. In this section, we consider the notion of the class of sub-exponential severity distribution models that may be utilized in OpRisk under an LDA framework to model risk processes which are suspected to contain potential loss events of significant monetary consequence.

One way to define an important family of distributions in OpRisk is through the notion of a heavy-tailed distribution. That is, one considers those distributions for which the moment-generating function does not exist on the positive real line such that

$$\int \exp(tx)dF(x) = \infty, \ \forall t > 0. \tag{3.25}$$

In other words, the standard Markov inequality for the exponentially decaying 'light' tail behaviour of a loss distribution in which

$$\overline{F}(x) \leq \exp(-sx)\mathbb{E}[\exp(sX)], \ \ \forall x > 0 \tag{3.26}$$

does not apply.

EXAMPLE 3.3 Examples of Heavy-Tailed Loss Distributions

The following are some relevant OpRisk loss distribution models that are heavy tailed, as characterized by the tail function of the loss distribution $\overline{F}(x) = 1 - F(x)$:

1. Burr distribution which has a tail function given by

$$\overline{F}(x) = \left(\frac{\beta}{x^{\gamma} + \beta}\right)^{\alpha}, \ \ x \in [0, \infty), \tag{3.27}$$

   with positive shape and scale parameters $\alpha, \beta, \gamma > 0$.

2. Pareto distribution which is basically a special case of the above Burr family, when $\gamma = 1$, and so it has a tail function given by

$$\overline{F}(x) = \left(\frac{\beta}{x + \beta}\right)^{\alpha}, \ \ x \in [0, \infty). \tag{3.28}$$

■

It can be shown that a distribution F has a moment-generating function in some right neighbourhood of the origin if and only if the following bound holds for some positive real numbers M and t,

$$\overline{F}(x) \leq M \exp(-tx), \ \forall x > 0. \tag{3.29}$$

Hence, one basic definition of an important class of distributions in OpRisk is the heavy-tailed distributions which have a right tail heavier than any exponential distribution. However, there are numerous more refined categorizations of heavy-tailed distributions which are required for the results in this chapter, see a tutorial review in Peters *et al.* (2013).

It will also be of relevance throughout this chapter and the text to utilize the following definition of the hazard function $R(x)$ and the hazard rate $r(x)$, see Definition 3.9.

Definition 3.9 (Hazard Function and Hazard Rate) *For a loss distribution F with support on $\mathbb{R}^+$, the hazard function is given by*

$$R(x) = -\ln \overline{F}(x). \tag{3.30}$$

If the loss distribution F is absolutely continuous with loss density f with respect to Lebesgue measure, then such a distributions tail behaviour can be characterized also by the hazard rate, given by

$$r(x) := \frac{dR(x)}{dx} = \frac{f(x)}{1 - F(x)} = \frac{f(x)}{\overline{F}(x)}. \tag{3.31}$$

In fact, in the case that the distribution F is a loss distribution on the positive real line, one can obtain the representation of the tail function of the loss distribution according to the hazard rate, given by

$$\overline{F}(x) = \exp\left(-\int_{-\infty}^{x} r(y)\,dy\right). \tag{3.32}$$

We observe from this definition that when we aim to characterize heavy-tailed distributions by their hazard functions, we are effectively studying the representation of the tails of the distribution according to the exponentiation of the negative of a monotone concave function.

Remark 3.6 (Hazard Rate of a Loss Distribution) *The following three right limiting possibilities of the hazard rate function of a loss distribution characterize its tail behaviour:*

1. *If the following limit of the hazard rate function $\lim_{x\to\infty} r(x) = 0$ holds, then the loss distribution F will be a heavy-tailed distribution function. Furthermore, one can show that if F is heavy tailed, then $\overline{F}$ is a heavy-tailed function, that is, decreases more slowly than an exponential such that*

$$\text{limsup}_{x\to\infty} \overline{F}(x)\exp(\lambda x) = \infty, \quad \forall \lambda > 0, \tag{3.33}$$

and in addition, if the density $f(x)$ exists, then it will also be a heavy-tailed function satisfying

$$\text{limsup}_{x\to\infty} f(x)\exp(\lambda x) = \infty, \quad \forall \lambda > 0. \tag{3.34}$$

Furthermore, one can show that in such a case, the hazard function $R(x)$ satisfies

$$\text{liminf}_{x\to\infty} \frac{R(x)}{x} = 0. \tag{3.35}$$

Conversely, it is also worth noting that in general, when the density $f(x)$ of a loss random variable is a heavy-tailed function, the distribution function $F(x)$ need not be a heavy-tailed function.

2. *If the following limit of the hazard rate function $\lim_{x\to\infty} r(x) > 0$ is satisfied, then the loss distribution F is not heavy tailed and the exponential moments will exist up to $\text{liminf}_{x\to\infty} r(x) > \lambda > 0$.*

3. *If the following limit of the hazard rate function* $\lim_{x\to\infty} r(x)$ *does not exist but one has* $liminf_{x\to\infty} r(x) = 0$*, then the distribution* F *can be either heavy or light tailed, and one needs further information to determine the characteristics.*

Foss *et al.* (2011, theorem 2.9) introduce the above concepts regarding characterization of the tail features of a loss distribution through the result in Theorem 3.2.

Theorem 3.2 (Generalized Moments to Characterize Heavy-Tailed Loss Models) *Consider a loss random variable* $X \sim F(x)$ *with heavy-tailed loss distribution taking positive support. Furthermore, consider a function* $g(x)$ *such that* $\lim_{x\to\infty} g(x) \to \infty$*. Then there exists a monotone, concave function denoted by* $h(x) : \mathbb{R}^+ \mapsto \mathbb{R}^+$ *such that* $h(x) = o(x)$ *as* $x \to \infty$ *with*

$$\mathbb{E}[\exp(h(X))] < \infty,$$
$$\mathbb{E}[\exp(h(X) + g(X))] = \infty.$$

We first discuss the class of sub-exponential distributions. Sub-exponential models arose as a consequence of the work on branching processes with age dependence in Chistyakov (1964), where asymptotic properties of such stochastic processes were studied. Several important results relating to convolutional properties of such classes of random variable were derived and are summarized in the following sections of this chapter. Before presenting such results, we first define the notion of the n-fold convolution to obtain distributions for linear combinations of n identically distributed independent random variables. Note, generally in OpRisk, we will have the identical distribution assumption; however, if this is not the case, one can trivially adjust the following definition to account for the sequence of distributions.

Definition 3.10 (n-Fold Convolution) *The distribution of the linear combination of independent identically distributed random variables* $X_1, \ldots, X_n$ *is the n-fold convolution* $F^{(n)*}(x) = F^{(n-1)*}(x) \star F(x)$*, where for any two distributions, one has*

$$F^{(2)*}(x) = (F \star F)(x) = F(x) \star F(x) = \int_{-\infty}^{\infty} F(x-y)\,dF(y). \tag{3.36}$$

which results in the recursive specification given by,

$$F^{(n)*}(x) = F^{(n-1)*}(x) \star F(x) = \int_{-\infty}^{\infty} F^{(n-1)*}(x-y)\,dF(y). \tag{3.37}$$

Furthermore, one can show that in the case that the loss random variables are not identically distributed, just independent, the following result applies to tighten the bound in Theorem 3.1 as shown in Theorem 3.3.

Theorem 3.3 (Bounding Probabilities of Convolutions of Independent Losses) *Consider two losses* $X_i \sim F_i(x)$*,* $i \in \{1, 2\}$ *each with support* $\mathbb{R}^+$ *and such that* X_1 *and* X_2 *are independent. Then one has the following result as long as either* F_1 *or* F_2 *or both are heavy-tailed severity distributions, resulting in*

$$\text{liminf}_{x\to\infty} \frac{\overline{F_1 \star F_2}(x)}{\overline{F}_1(x) + \overline{F}_2(x)} = 1. \tag{3.38}$$

To illustrate this concept, we can consider the following simple convolution results for the linear combination of two independent random variables given in Example 3.4.

EXAMPLE 3.4

The density and distribution of the sum of two independent continuous random variables $X_1 \sim F_1(\cdot)$ and $X_2 \sim F_2(\cdot)$ with densities $f_1(\cdot)$ and $f_2(\cdot)$ respectively, can be calculated via convolution as

$$f_{X_1+X_2}(x) = (f_1 \star f_2)(x) = \int f_2(x - x_1) f_1(x_1) dx_1,$$

and

$$\begin{aligned} F_{X_1+X_2}(x) &= (F_1 \star F_2)(x) \\ &= \int F_2(x - x_1) dF_1(x_1) \\ &= \int F_2(x - x_1) f_1(x_1) dx_1, \end{aligned}$$

respectively. Consider the example in which X_1 and X_2 are independent losses with inverse Gaussian distributions,

$$\begin{aligned} X_1 &\sim InverseGaussian(\mu_0 w_1, \lambda_0 w_1^2), \\ X_2 &\sim InverseGaussian(\mu_0 w_2, \lambda_0 w_2^2). \end{aligned}$$

Then, the sum of the two losses has a known density and distribution given as the solution to the convolution integrals as

$$\begin{aligned} f_{X_1+X_2}(x) &= \left[\frac{\lambda_2}{2\pi x^3}\right]^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu_2)^2}{2\mu_2^2 x}\right) \\ F_{X_1+X_2}(x) &= \Phi\left(\sqrt{\frac{\lambda_2}{x}}\left(\frac{x}{\mu_2} - 1\right)\right) \\ &\quad + \exp\left(\frac{2\lambda_2}{\mu_2}\right)\Phi\left(-\sqrt{\frac{\lambda_2}{x}}\left(\frac{x}{\mu_2} + 1\right)\right), \end{aligned}$$

with

$$\begin{aligned} \mu_2 &= \mu_0 \sum_{i=1}^{2} w_i, \\ \lambda_2 &= \lambda_0 \left(\sum_{i=1}^{2} w_i\right)^2. \end{aligned}$$

■

Having defined the integral equations required to be solved to obtain the distribution of linear combinations of independent random variables, one may now present the notion of a sub-exponential distribution, which will be represented by the class of

distributions, $F \in \mathscr{F}$. First, we recall the definition of the important class of functions, which we will denote by $\mathscr{F}_1(\gamma)$, detailed in Definition 3.11 and given in the context of branching phenomena for positive random variables in Chistyakov (1964) and Chover *et al.* (1973) and then later in the insurance context in Embrechts and Goldie (1982).

Definition 3.11 *A distribution F with support $[0, \infty[$ belongs to the class $F \in \mathscr{F}_1(\gamma)$ for $\gamma \geq 0$ iff it satisfies the following properties:*

1. *Tail ratios under translation satisfy asymptotically the relationship*

$$\lim_{x \to \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = \exp(\gamma y), \quad \forall y \in \mathbb{R}. \tag{3.39}$$

2. *Convolution tail ratios are finite for second-order convolution asymptotically*

$$\lim_{x \to \infty} \frac{\overline{F}^{(2)*}(x)}{\overline{F}(x)} = c < \infty. \tag{3.40}$$

3. *First derivative (density) evaluated at γ satisfies*

$$f(-\gamma) = \int_0^\infty \exp(\gamma x)\, dF(x) < \infty. \tag{3.41}$$

A second class, denoted here by $\mathscr{F}_2(\gamma)$ and given in Theorem 3.4, was also studied in Teugels (1975) and Klüppelberg (1989), this class is identical to the one above in Definition 3.11 only in the special case of $\gamma = 0$. To show that a distribution is a member of this class $\mathscr{F}_2(\gamma)$ for some $\gamma \geq 0$, it was shown in Embrechts and Goldie (1982) that one needs only to verify the conditions in Theorem 3.4.

Theorem 3.4 *A distribution F with support $[0, \infty[$ is a member of the class of functions $\mathscr{F}_2(\gamma)$ for some $\gamma \geq 0$ iff it satisfies:*

1. *The density satisfies $f(-\gamma) < \infty$; and*
2. *The γ-Transform given by*

$$F(x;\gamma) = \frac{1}{f(\gamma)} \int_0^x \exp(\gamma y)\, dF(y) \tag{3.42}$$

satisfies that $F(x;\gamma) \in \mathscr{F}(0)$.

As discussed in Klüppelberg (1989), the family of sub-exponential distributions $\mathscr{F}(\gamma)$ is closed with respect to tail equivalence. This property ensures that any distribution in the class $\mathscr{F}(\gamma)$ can always be assumed to be absolutely continuous. Recall that two distributions F and G are tail equivalent if there exists a $c \in (0, \infty)$ such that

$$\lim_{x \to \infty} \frac{\overline{F}(x)}{\overline{G}(x)} = c. \tag{3.43}$$

Remark 3.7 *In the special case that $\gamma = 0$, these two classes $\mathscr{F}_1(0)$ and $\mathscr{F}_2(0)$ of distributions coincide and are known as the sub-exponential family, denoted by $\mathscr{F}$. However, it was proven in Embrechts and Goldie (1982) that if $\gamma > 0$, then the two classes do not coincide any longer, $\mathscr{F}_1(\gamma) \neq \mathscr{F}_2(\gamma)$.*

More precisely, the definition of the class of heavy-tailed models given by the sub-exponential family of severity distributions that we denote by membership $(F(x) \in \mathscr{F})$ is provided in Definition 3.11, see further discussion in, for example, Resnick (2006) and Klüppelberg (1989).

Definition 3.12 (Sub-exponential Severity Models) *The sub-exponential family of distributions $\mathscr{F}$ defines a class of heavy-tailed severity models that satisfy the limits*

$$\lim_{x\to\infty} \frac{1 - F^{(n)*}(x)}{1 - F(x)} = n, \tag{3.44}$$

if and only if

$$\lim_{x\to\infty} \frac{1 - F^{(2)*}(x)}{1 - F(x)} = 2. \tag{3.45}$$

To understand this condition of sub-exponentially, consider the setting in OpRisk of a loss distribution with positive support for the linear combination of two i.i.d. loss random variables. Then for the total loss, one would be considering the two-fold convolution of the tail

$$\begin{aligned}
\overline{F}^{(2)*}(x) &= \mathbb{P}\mathrm{r}[X_1 + X_2 \geq x] \\
&\geq \mathbb{P}\mathrm{r}[\{X_1 \geq x\} \cup \{X_x \geq x\}] \\
&= \mathbb{P}\mathrm{r}[X_1 \geq x] + \mathbb{P}\mathrm{r}[X_2 \geq x] - \mathbb{P}\mathrm{r}[X_1 \geq x, X_2 \geq x] \\
&= 2\overline{F}(x) - \overline{F}(x)^2 \\
&= 2\overline{F}(x)(1 + o(1)).
\end{aligned} \tag{3.46}$$

Then as $x \to \infty$ one has from sub-exponentiality the property

$$\mathrm{limsup}_{x\to\infty} \frac{\overline{F}^{(2)*}(x)}{\overline{F}(x)} \leq 2. \tag{3.47}$$

One can also show that the sub-exponential distribution properties satisfy the following relationship for i.i.d. losses $\{X_i\}_{i=1}^n$ aggregated into a partial sum which gives, via the definition of a sub-exponential distribution, the result asymptotically as $x \to \infty$ given by

$$\overline{F}_{Z_n}(x) = \mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x] \sim n\mathbb{P}\mathrm{r}[X_1 > x].$$

In addition, one can consider the distribution of the maximum

$$\begin{aligned}
\mathbb{P}\mathrm{r}[\max\{X_1 + \cdots + X_n\} > x] &= \overline{F}_X(x)^n \\
&= \overline{F}_X(x)(1 - F_X(x))^{n-1} \\
&= \overline{F}_X(x)(1 + F_X(x) + \cdots + F_X(x)^{n-1}) \\
&\sim n\overline{F}_X(x), \quad as \quad x \to \infty.
\end{aligned}$$

Hence, one may conclude that a sub-exponential distribution satisfies the property that the two following tail probabilities are asymptotically ($x \to \infty$) equivalent, with

$$\Pr[X_1 + \cdots + X_n > x] \sim \Pr[\max\{X_1 + \cdots + X_n\} > x].$$

This property implies that the partial sum is likely to get large when one of the loss r.v.'s gets large. This sub-exponential heavy-tail feature arising from a so-called 'single big jump' or 'one loss causes ruin' situation can be understood through the following example, see discussion in Hallinan Jr (1993) and Foss *et al.* (2013).

EXAMPLE 3.5 Weibull - Stretched Exponential Severity Model

Consider a loss distribution F given by the widely used Weibull distribution with density given by

$$f(x;\gamma) = \gamma x^{\gamma-1}\exp(-x^{\gamma}), \quad x \geq 0, \tag{3.48}$$

for a shape parameter $\gamma > 0$ and with a tail function given by

$$\overline{F}(x;\gamma) = \exp(-x^{\gamma}), \quad x \geq 0. \tag{3.49}$$

This loss distribution will possess the characteristics of a heavy-tailed model when $\gamma \in (0,1)$. Now consider the distribution of two losses X_1 and X_2 each i.i.d. from this Weibull loss distribution when $\gamma < 1$. Then the probability density for $c^{-1}X_1 \leq x$ conditional upon the event $X_1 + X_2 = c$, for some $c > 0$, is given up to proportionality by

$$f_{c^{-1}X_1|X_1+X_2=c}(x;\gamma) \propto \exp(-c^{\gamma}(x^{\gamma} + (1-x)^{\gamma}))[x(1-x)]^{\gamma-1}. \tag{3.50}$$

In the case that $\gamma < 1$, in which case one has a heavy-tailed loss model, then the mass of the corresponding distribution, obtained by integrating this density between $[0, x]$, as $c \to \infty$ concentrates more and more near the origin and also on the extreme loss boundary characterized by $X_1 + X_2 = c$. This provides a simple intuitive example of the one loss causes ruin framework. ■

The interpretation of the notion of a sub-exponential distribution then comes from considering the fact that for distributions that satisfy $F \in \mathscr{F}$, one has that

$$\lim_{x\to\infty} \exp(\varepsilon x)\overline{F}(x) = \infty, \quad \varepsilon > 0. \tag{3.51}$$

This clearly justifies the name given to this class because it shows that the tail of the distribution $\overline{F}(x)$ will tend to zero slower than any exponential $\exp(-\varepsilon x)$.

One important property of sub-exponential random variables is that given n i.i.d. loss random variables $X_i \sim F \in \mathscr{F}$ for $i \in \{1, 2, \ldots, n\}$, they will be max-sum equivalent, which means that defining the distribution $\max\{X_1, X_2, \ldots, X_n\} \sim G$ and the distribution $\sum_{i=1}^{n} X_i \sim F_{Z_n}$ then max-sum equivalence means that

$$\overline{G} \sim \overline{F}_{Z_n} \sim \sum_{i=1}^{n} \overline{F}_{X_i}, \quad x \to \infty. \tag{3.52}$$

To understand a further characterization of the sub-exponential family, one considers the notion of an asymptotically locally constant function given in Definition 3.13.

Definition 3.13 (Asymptotically Locally Constant Distributions) *A loss distribution $F(x) > 0$ is said to be asymptotically locally constant if it satisfies for any fixed constant $y \in [0, \infty)$ the condition*

$$\lim_{x \to \infty} \frac{F(x+y)}{F(x)} \to 1. \tag{3.53}$$

An analogous definition of such a condition relates to what is known as *local sub-exponentiality*, given in Definition 3.14.

Definition 3.14 (Local Sub-exponentiality) *A distribution F is said to be locally sub-exponential, denoted by $F \in \mathscr{F}_{\Delta}$ if the distribution satisfies two conditions:*

1. *Firstly, that $F(x + y + \Delta) \sim F(x + \Delta)$ holds uniformly in $|y| \leq 1$ for some $\Delta = (0, T]$ with $T \in (0, \infty]$; and*
2. *Secondly, that for $F_0(x) = \frac{F(x)}{F(\infty)}$, one has that $F_0^{(2)*}(x + \Delta) \sim 2F_0(x + \Delta)$.*

The property of asymptotic local constant distributions was utilized in Pitman (1980) to demonstrate that the necessary and sufficient condition for membership of a severity distribution in the sub-exponential class ($F \in \mathscr{F}$) is satisfied if and only if the tail distribution $\overline{F}(x) = 1 - F(x)$ satisfies

$$\lim_{x \to \infty} \int_0^x \frac{\overline{F}(x-y)}{\overline{F}(x)} F(y)\, dy = 1.$$

Alternatively, one may characterize the family of distributions $F \in \mathscr{F}$ by those that satisfy asymptotically the tail ratio

$$\lim_{x \to \infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1, \ \forall y \in [0, \infty). \tag{3.54}$$

Severity models $F \in \mathscr{F}$ are of interest for severity distributions in high consequence loss modeling because they include models with *infinite mean loss* and *infinite variance*. In addition, the class $\mathscr{F}$ includes all severity models in which the tail distribution under the log transformed r.v., $\overline{F}(\ln(x))$, is a slowly varying function of x at infinity. We will discuss both regular and slow variation later.

Examples of models in this family include:

1. Pareto: $\overline{F}(x) = \left(\frac{c}{c+x}\right)^{\alpha}$ for $x \geq 0$, $c > 0$ and $\alpha > 0$.
2. LogNormal: $\overline{F}(x) = \frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right]$, $x \geq 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$.
3. Heavy-tailed Weibull: $\overline{F}(x) = \exp(-\lambda x^{\alpha})$, $x \geq 0$, $\lambda > 0$ and $0 < \alpha < 1$.

An important characteristic property of sub-exponential distributions is their asymptotic tail additivity of the convolutions of the original distributions. For instance, if one has loss random variables which are i.i.d. with $X_i \sim F(x) \in \mathscr{F}$, then the tail of the linear combination of these loss random variables distribution will asymptotically look like the linear combination of the distribution tails as characterized by the tail additivity property

$$\mathbb{P}\mathrm{r}[X_1 + X_2 + \cdots + X_n \geq x] \sim n\mathbb{P}\mathrm{r}[X_1 \geq x], \tag{3.55}$$

which is effectively a restatement of the max-sum equivalence property discussed above. In OpRisk, one is typically interested in the right tail of the loss model; fortunately, for positive-supported distributions, sub-exponential characterizations of heavy-tailed loss models is ideal as it is by definition a property of right tails of a loss distribution only. In relatively recent work there has been some efforts to extend the notion of sub-exponential families to the entire real line, see further details in Foss *et al.* (2013).

3.3 Introduction to the Regular and Slow Variation Families of Heavy-Tailed Models

Before proceeding with further properties of distribution functions that are members of the class of sub-exponential severity models $\mathscr{F}$, it will be useful to first recall the notions of variation of a distribution function in the tails (very high or very low quantiles). This naturally leads one to consider the definition of the concept of a slowly varying function (SV) and regularly varying (RV) functions along with some relevant properties. Having, defined these concepts, we will then be able to define a second larger class of distributions, denoted by $\mathscr{L}$, that we will then relate to the sub-exponential class and will be of significance in the analysis to follow.

Leading on from the specification of the sub-exponential family, one may also consider classes of heavy-tailed severity distributions as classified by their right tail properties through formal definitions such as regularly varying tail, long-tailed, dominantly varying tail, subversively varying tail and smoothly varying tail, each of which we briefly define later. We then discuss the relationship that these different classes of severity model assumptions have with each other to provide a basic understanding of the relationships between each of these possible heavy-tailed severity modeling assumptions.

Arguably, one of the most utilized sub-classes of the sub-exponential distributions is the class of regularly varying distributions. The definition of the class of regularly varying functions given by Definition 3.15, see Bingham *et al.* (1989) and Resnick (2006).

Definition 3.15 (Regular Variation) *A measurable function $f(x) > 0$ that satisfies the condition that*

$$\lim_{x\to\infty} \frac{f(\lambda x)}{f(x)} \to \lambda^{\rho},\ \forall \lambda > 0 \tag{3.56}$$

is regularly varying with index ρ denoted by $f \in RV_{\rho}$. That is, the function satisfies the asymptotic equivalence $f(\lambda x) \sim \lambda^{\rho} f(x)$ as $x \to \infty$.

Remark 3.8 *We note that when referring to a loss distribution F as regularly varying at infinity with index $-\rho < 0$, this means that the tail function has the following asymptotic equivalence, $\overline{F}(\lambda x) \sim \lambda^{-\rho}\overline{F}(x)$.*

We define the class of all regularly varying functions denoted by $\mathscr{R} = \cup_{\rho\in\,\mathbb{R}} RV_{\rho}$. It is also convention to distinguish the special sub-class of functions denoted generically by $L(x)$ that are regularly varying with an index of $\rho = 0$ as follows.

Corollary 3.2 (Slowly Varying Tail) *A function $f \in \mathscr{R}$ is slowly varying if* $\rho = 0$.

So one may consider the class of severity distributions or densities that are members of this class. To provide some intuition for properties of severity models in this class of regularly varying models, we note the following important features. It can be shown that for loss processes in which the severity models have a strictly non-negative support, the membership of a severity distribution in the class of regularly varying functions with tail index $\rho > 0$ implies that

$$\mathbb{E}\left[X^{\beta}\right] = \begin{cases} c, & \text{for some real constant } c < \infty, \text{if } \beta < \rho; \\ \infty, & \text{if } \beta > \rho. \end{cases} \tag{3.57}$$

Analogously (setting $\beta = \rho + k$), if the right tail of a distribution $\overline{F}$ is regularly varying with an index ρ, then this implies that the distribution F will have $(\rho + k)$-th moments which are infinite for $k > 0$. In addition, the following is also known about truncated moments of severity distributions in which $\overline{F}_X(x)$ is regularly varying at infinity with $\rho = -\alpha$ for some $\alpha > 0$.

$$\begin{aligned} \mathbb{E}\left[X^{\beta}\mathbb{I}_{X \le x}\right] &\sim \frac{1}{\beta - \alpha} x^{\beta}\, \overline{F}_X(x), \quad \beta > \alpha. \\ \mathbb{E}\left[X^{\beta}\mathbb{I}_{X > x}\right] &\sim \frac{1}{\beta - \alpha} x^{\beta}\, \overline{F}_X(x), \quad \beta \le \alpha. \end{aligned} \tag{3.58}$$

Clearly, these truncated moments are often of direct interest in risk management and especially in the context of capital estimation, see Peters *et al.* (2013).

In addition, following properties of regularly varying functions apply specifically in the context of distributions and densities, see Soulier (2009, theorem 1.20) and Mikosch (1999).

Corollary 3.3 (Regularly Varying Severity Distributions) *If the severity model has a regularly varying distribution $F \in RV_{\rho}$ with a density f which is locally integrable on $[1, \infty)$ with*

$$F(x) = \int_1^x f(t)\,dt, \tag{3.59}$$

then given the severity density f is ultimately monotone, one has for $\rho \neq 0$ that $f \in RV_{\rho-1}$. Furthermore, one can show that in the case of non-negative random variables, such as for a loss process in an LDA severity model, if the distribution is regularly varying $F \in RV_{\rho}$ with $\rho \geq 0$, then the right tail $\overline{F}(x) \in RV_{-\rho}$.

Hence, if we apply the above results and definitions in the OpRisk context, consider the severity loss distribution $F(x)$. If the right tail of a distribution $\overline{F}$ is regularly varying with an index ρ, then this means that the asymptotic limit $(x \to \infty)$, for all $\lambda > 0$, satisfies the relationship

$$\lim_{x \to \infty} \frac{\overline{F}(x\lambda)}{\overline{F}(x)} = \lambda^{-\rho}. \tag{3.60}$$

The implications of this for this OpRisk context show the heavy-tailed behaviour of such a class of distributions because it explains why the distribution F will have $(\rho + k)$-th moments which are infinite for $k > 0$.

In addition, one may also present several other fundamental properties of slowly varying and regularly varying functions that are known, see, for instance, Borovkov and Borovkov (2002, theorem 1.1.4), from which we select a few of these properties that will be of relevance in heavy-tailed severity models in OpRisk settings to follow in future chapters. In Proposition 3.1, we provide some of the particularly utilized properties of functions in the class RV_0, defining functions of slow variation.

Proposition 3.1 (Properties of Slow and Regular Varying Functions) *Consider two slowly varying functions given by* $L(x) = L_1(ax + b)$ *and* $L_2(x)$ *and some constants* $a \geq 0$ *and* $b \in \mathbb{R}$. *The following results provide properties of functions in the class of regular and slow variation:*

- *The function operations* $L_1 + L_2$ *and* $L_1 L_2$ *will produce new functions which are also slowly varying functions, that is, the class of slowly varying functions* RV_0 *is closed under the operations of addition, multiplication and division.*
- *If a function* $L \in RV_0$ *then* $L^b \in RV_0$ *for all* $b \in \mathbb{R}$.
- *Given an index of regular variation* $\beta \in \mathbb{R}$, *then* $f \in RV_\beta$ *if and only if* $f^{-1} \in RV_{-\beta}$.
- *If a function* $f \in RV_\beta$, *with* $\beta > 0$, *then* $\lim_{x \uparrow \infty} f(x) = \infty$ *and, with* $\beta < 0$, *then* $\lim_{x \uparrow \infty} f(x) = 0$.
- *If a function* $L \in RV_0$, *then*

$$\lim_{x \uparrow \infty} \frac{\ln L(x)}{\ln x} = 0. \tag{3.61}$$

- *If a function* $L \in RV_0$, *then for each* $\gamma > 0$, *there exists a value* x_γ *so that for all constants* $a > 0$ *and* $x > x_\gamma$ *the following bound holds*

$$ax^{-\gamma} < L(x) < ax^{\gamma}. \tag{3.62}$$

It is worthwhile considering through some Examples 3.6 what functional forms some members of RV_0 will have.

EXAMPLE 3.6

Examples of functions $L \in RV_0$ include the following:

1. It can be shown that $L(x) = \ln(x)^\alpha$ for any $\alpha \in \mathbb{R}$ such that $L \in RV_0$. Note it is understood that the log function is only considered asymptotically as $x \to \infty$.
2. It can be shown that $L(x) \to C \in (0, \infty)$ for some constant C provides a function $L \in RV_0$.
3. It can be shown that $L(x) = \exp\{\ln(x)^\beta\}$ for any $\beta < 1$ will produce a function such that $L \in RV_0$.

■

Typical examples of slowly varying functions include:

1. Functions converging to a positive constant, logarithms and iterated logarithms;
2. Distributions such as the Pareto, Cauchy, Student-t, Burr and log-gamma truncated α-stable distributions have regularly varying tails and are ultimately infinitely differentiable and their derivatives are regularly varying.

Having presented several properties of regular and slowly varying functions, it is also important to consider two important and highly utilized properties of such functions, namely, the uniform convergence property and the integral representation they admit. First, we introduce the background of these concepts and then present their formal definitions and implications for modeling. We begin by recalling the notion of pointwise convergence and uniform convergence of a sequence, given in Definitions 3.16.

Definition 3.16 *A sequence of functions* $\{f_n(x)\}$ *with domain* $\mathscr{D}$ *converges pointwise if for each fixed* $x_0 \in \mathscr{D}$*, the numeric sequence* $\{f_n(x_0)\}$ *converges as a numeric sequence. That is, for each fixed* x_0 *and any given* $\varepsilon > 0$*, there exists a positive integer* N *such that*

$$|f_n(x_0) - L| < \varepsilon, \ \forall n \geq N, \tag{3.63}$$

for some limit L *which depends on* x_0 *and* N *depends on* x_0 *and* ε*. Furthermore, if* $\{f_n(x)\}$ *is converging pointwise, then* $\lim_{n\to\infty} f_n(x) = f(x)$ *is a well-defined function.*

A stronger form of convergence, known as *uniform convergence*, can also be defined in which the convergence now holds with the integer N only a function of ε as given in Definition 3.17.

Definition 3.17 *A sequence of functions* $\{f_n(x)\}$ *with domain* $\mathscr{D}$ *converges pointwise uniformly if for each fixed* $x_0 \in \mathscr{D}$*, the numeric sequence* $\{f_n(x_0)\}$ *converges as a numeric sequence. That is, for each fixed* x_0 *and any given* $\varepsilon > 0$*, there exists a positive integer* N *such that*

$$|f_n(x_0) - L| < \varepsilon, \ \forall n \geq N, \tag{3.64}$$

for some limit L *which depends on* x_0 *and* N *depends only on* ε*.*

Now given the definition of uniform convergence, one can show that for a function of slow variation, the rate of convergence is uniform over its domain, as given in Theorem 3.5 which will lead us to the Karamata characterization and the Representation Theorem. The uniform convergence property is given in Theorem 3.5, see Borovkov and Borovkov (2002). This result shows that the uniform convergence of the slowly varying function will occur on an interval $[\gamma^{-1}, \gamma]$ as $x \to \infty$ as long as $\gamma = \gamma(x)$ increases at a slow enough rate.

Theorem 3.5 (Uniform Convergence of Regularly Varying Functions) *If* $L(x)$ *is a slowly varying function as* $x \to \infty$*, then convergence given by the ratio*

$$\lim_{x\to\infty} \frac{L(x\gamma)}{L(x)} \to 1 \tag{3.65}$$

will hold uniformly in $\gamma \in [\gamma_1, \gamma_2]$ *for some* γ_1 *and* γ_2 *satisfying* $0 < \gamma_1 < \gamma_2 < \infty$*.*

Furthermore, in OpRisk, when considering severity densities and distribution functions we are working with strictly positive functions. In this regard, we note that positive regularly varying functions have a unique representation detailed in Theorem 3.6, see Bojanic and Karamata (1963), Bojanic and Seneta (1971), Geluk (1981) and Balkema *et al.* (1979). This representation demonstrates an important property of such positive regularly varying functions, as it shows that the integration of regularly varying functions (tail functionals) will behave in the same manner as the integration of power functions. In addition, we note that, in general, the class of severity distributions and densities considered in OpRisk settings will be strictly

monotonic in their tail behaviour. This is significant as it means that one will achieve uniform convergence in the limit taken in the definition of regular variation for such severity distributions.

Theorem 3.6 (Karamata's Representation Theorem for Regularly Varying Functions) *A function f is a positive regularly varying function at infinity, $f \in RV_{-\rho}$, with index $-\rho$ if and only if f can be represented by*

$$f(x) = c(x) \exp\left(\int_{x_0}^{x} \frac{-\rho + \varepsilon(t)}{t} dt \right), \ \forall x \geq x_0, \tag{3.66}$$

with $c(x) = c + o(1)$ for some $c > 0$ and $\varepsilon(t) = o(1)$.

Remark 3.9 *It is worth considering the intuition and relevance of the Karamata Representation Theorem. In particular, it demonstrates that when integrating regularly varying functions $f \in RV_{\rho}$, one can pass the slowly varying component outside the integral as follows*

$$F(x) = \int_0^x f(t)dt = \int_0^x t^{\rho} L(t)dt = L(x) \int_0^x t^{\rho} dt = \frac{x^{\rho+1} f(x)}{\rho + 1}, \tag{3.67}$$

see discussion in Resnick (2006, p. 25).

Another way of seeing this is that if $\rho > 1$, then for a function $f \in RV_{\rho}$ one can consider the asymptotic equivalence given by

$$\int_x^{\infty} f(u)du \sim \frac{xf(x)}{\rho - 1}, \quad as \ \ x \to \infty, \tag{3.68}$$

and if $\rho < 1$, then one can consider the asymptotic equivalence given by

$$\int_0^x f(u)du \sim \frac{xf(x)}{1 - \rho}, \quad as \ \ x \to \infty \tag{3.69}$$

and, finally, when $\rho = 1$, one obtains

$$\int_0^x f(u)du = xf(x)L_1(x), \quad as \ \ x \to \infty \tag{3.70}$$

for $L_1(x) \in RV_1$.

Remark 3.10 *A consequence of this representation theorem is that every regularly varying function $f \in RV_{\rho}$ will admit a representation given by*

$$f(x) = x^{\rho} L(x). \tag{3.71}$$

Remark 3.11 *One way to understand the implications of working with severity models from this class of distributions is to consider what it tells the practitioner about integration of functions with respect to such severity distributions. Of course, the key component in this regard will be the influence the right tail plays in such functionals. In particular, if a severity density $f \in RV_{\rho}$, then if $\rho \geq -1$, the integration of such a function satisfies*

$$1 - \overline{F}(x) = \int_0^x f(t)dt \in RV_{\rho+1}. \tag{3.72}$$

Alternatively, if the index of regular variation is $\rho \leq -1$*, then one has*

$$\overline{F}(x) = \int_0^x f(t)dt \in RV_{\rho+1}. \tag{3.73}$$

In addition, a distribution F which is regularly varying with index ρ can be characterized by what is widely known as the *tail balance condition* in Definition 3.3, see Jessen and Mikosch (2006).

Lemma 3.3 (Tail Balance Condition) *A distribution function* F *is regularly varying with index* $\rho \geq 0$ *if there exists* $p, q \geq 0$ *with* $p + q = 1$ *and a slowly varying function* $L(x)$*, which for all* $\lambda > 0$ *means*

$$\lim_{x\to\infty} \frac{L(x\lambda)}{L(x)} = 1, \tag{3.74}$$

and the following tail balance conditions are satisfied as $x \to \infty$*,*

$$\begin{aligned} F(-x) &= qx^{-\rho}L(x)(1 + o(1)), \\ \overline{F}(x) &= px^{-\rho}L(x)(1 + o(1)). \end{aligned} \tag{3.75}$$

One can also obtain the following bounds using properties of functions of slow variation $L(x) \in \mathscr{R}_0$. Below, we present a set of bounds for ratios of such functions evaluated pointwise as given by the Potter Bounds in Proposition 3.2.

Proposition 3.2 (Potter Bounds) *Given a slowly varying function* $l \in RV_0$ *evaluated at two points in its domain* $x, y \in \mathscr{D}(l)$ *such that* $x, y > x_0$ *and a constant* $a > 1$ *and* $\gamma > 0$*, then there exists a constant* $x_0(a, \gamma)$ *(dependent on* a *and* γ*) such that the following bound is satisfied*

$$\frac{L(y)}{L(x)} \leq a \max\left\{ \left(\frac{y}{x}\right)^{\gamma}, \left(\frac{y}{x}\right)^{-\gamma} \right\}, \; x, y \geq x_0. \tag{3.76}$$

In OpRisk settings, it will also often be important when considering capital approximations to make the connection between the properties of the regularly varying distribution and its inverse (quantile function) that is often utilized pointwise as the mathematical measure of OpRisk capital, such as in the Basel II/III stipulated Value-at-Risk for some quantile level α. To see how knowledge of the regular variation of a distribution can aid in understanding the behaviour of the resulting quantile function one must first consider the definition of the De Bruijin (or De Bruyn) conjugate, Definition 3.18, for functions of slow variation.

Definition 3.18 (De Bruyn Conjugate) *If* L *varies slowly, there exists a slowly varying function, denoted by* $L^\sharp$*, which is unique up to an asymptotic equivalence, with the following conjugate (inverse) properties*

$$L(x)L^\sharp(xL(x)) \to 1, \quad L^\sharp(x)L(xL^\sharp(x)) \to 1, \tag{3.77}$$

as $x \to \infty$*, then* $[L^\sharp]^\sharp \sim L$*.*

See a proof of this result in Bingham *et al.* (1989, theorem 1.5.13).

Remark 3.12 *The consideration of a slowly varying function L and its de Bruijin conjugate $L^\sharp$ commonly occur in problems dealing with asymptotic inversion. The two applications of direct relevance to OpRisk practitioners of such a result regards: domain of attraction problems in probability and risk measure estimation such as single-loss approximations for compound processes.*

Proposition 3.3 (EVT and De Bruyn Conjugates) *A severity distribution F that satisfies (**Condition** D_I) in Definition 2.16 or the related conditions 2.3, which has a maximum limiting distribution $F \in \mathscr{D}(M)$ for which the EVI, $\gamma > 0$, can always be represented equivalently either in terms of the distribution function $F(x)$ or the tail quantile function $U(x)$ specified according to*

$$1 - F(x) = x^{-1/\gamma} L_F(x) \text{ and } U(x) = x^\gamma L_U(x), \tag{3.78}$$

for slowly varying functions $L_F(x)$ and its de Bruyn conjugate $L_U(x) = L_F^\sharp(x)$.

The proof of this result is provided in Beirlant *et al.* (2004, p. 79) and based on the de Bruyn conjugate for slowly varying functions. One can now utilize these ideas to consider what can be understood about the quantile function of a regularly varying loss distribution as stated in Lemma 3.7 and Definition 3.19.

Lemma 3.4 (Regularly Varying Distribution and Quantile Functions) *If the right tail of a distribution is regularly varying at infinity such that $\overline{F} \in RV_{-\rho}$ with $\rho > 0$ and $\overline{F}(x) = x^{-\rho} L(x)$, then it is also true that the quantile function $F^{\leftarrow}(t) = \inf\{x : F(x) \geq t\}$ will be regularly varying at the origin, $F^{\leftarrow}(t) \in RV_{-\rho}$. If one defines $Q(t) := F^{\leftarrow}(1 - 1/t)$ on $[1, \infty)$, then this leads to the representation $Q(t) = t^{\frac{1}{\rho}} L^\sharp\left(t^{1/\rho}\right)$, where $L^\sharp$ represents the De Bruyn conjugate of the composite function $L^{-1/\rho}$.*

EXAMPLE 3.7 G-and-H Severity Distribution

In this example, we consider the G-and-H family of severity distribution models characterized by the quantile transform of a base distribution such as a Gaussian. This is what is widely known in statistics as an elongation-based transform family of distributions. For instance, consider the G-and-H family characterized by transform of the Gaussian random variable $W \sim Normal(0, 1)$ given by

$$X \stackrel{d}{=} a + b \frac{\exp(gW) - 1}{g} \exp\left(\frac{hW^2}{2}\right) \tag{3.79}$$

with location a, scale b, skewness transform g and kurtosis transform h. Note, if $h = 0$, one obtains a scaled LogNormal distribution known as the *g-family of severity models*. In general, one can show that the tail behaviour of the G-and-H distribution with base distribution given by a Gaussian will have a distribution tail function that is regularly varying. The index of regular variation can be obtained by considering the result of the limit

$$\begin{aligned} &\lim_{x \to \infty} \frac{x f(x; a, b, g, h)}{1 - F(x; a, b, g, h)} \\ &= \lim_{x \to \infty} \frac{\phi(u)[\exp(gu) - 1]}{(1 - \Phi(u)[g \exp(gu) + hu(\exp(gu) - 1)]} \\ &= \frac{1}{h}, \end{aligned} \tag{3.80}$$

where $k(x) = g^{-1}(\exp(gx) - 1)\exp\left(hx^2/2\right)$ with $u = k^{-1}(x)$. This shows that the h-transform introduces a Pareto heavy-tailed feature to the transformed-based distribution tails. Because this shows that $\overline{F}(x) \in RV_{-1/h}$, then one has $\overline{F}(x) = x^{-1/h}L(x)$ as $x \to \infty$ where for the family of G-and-H models one can find the form of the slowly varying function given by

$$L(x) = \frac{1}{\sqrt{2\pi}g^{1/h}} \frac{\left[\exp\left(\frac{g}{h}\sqrt{g^2 + 2h\ln(gx)} - \frac{g^2}{h}\right) - 1\right]^{1/h}}{\frac{1}{h}\sqrt{g^2 + 2h\ln(gx)} - \frac{g}{h}} \left(1 + O\left(\frac{1}{\ln x}\right)\right). \tag{3.81}$$

One can show that this representation of the slowly varying function for the G-and-H family further indicates that this model belongs to the domain of attraction of an extreme value distribution with tail index h. ■

Definition 3.19 (Local Heavy Tailedness) *Given a distribution function $F(x)$ with a tail quantile function $U = (1/(1-F))^{\leftarrow}$ which satisfies $U \in RV_\beta$ and therefore that $\overline{F} \in RV_{-1/\beta}$, one can write $U(t) = \exp(\psi(\ln t))$ where one has for $t \to \infty$ that $\psi'(t) \to \beta$ and one interprets for fixed values of t the quantity $\psi'(\ln t)$ as the local tail index at t. That is, it is a measure of local heavy tailedness in which one can interpret the tail index β as the ultimate slope in the log–log plot of U.*

It will also be of interest in several applications when working with regularly varying distributions and densities for OpRisk modeling to consider what happens to such functions under certain classes of transformation such as a Laplace transform or Fourier transform with respect to a heavy-tailed loss random variable. Consider the distribution function $F(x) \in RV_\rho$ with density $f(x)$ and take the Laplace transform of the random variable with this distribution according to

$$\Psi(\theta) := \mathbb{E}[\exp(-\theta X)] = \int_0^\infty \exp(-\theta x) f(x)\, dx < \infty, \quad \text{for any } \theta > 0. \tag{3.82}$$

Then as discussed in Borovkov and Borovkov (2002, theorem 1.1.5), one can show the following inherited properties of a transformed regularly varying function under such transformations as given by the Laplace transform, see Theorem 3.7. The concepts presented later are related to the notion of a Tauberian theorem where instead of assuming a loss random variable has a distribution function which is of regular variation in its tail, instead one assumes it satisfies a monotonicity condition and then one assumes that the transform (Laplace) transform is of regular variation near the origin.

Theorem 3.7 (Regularly Varying Distributions Under Transformation) *Consider a distribution function $F(x) \in RV_\rho$ with density $f(x)$ and take the Laplace transform of the random variable with this distribution according to*

$$\Psi(\theta) := \mathbb{E}[\exp(-\theta X)] = \int_0^\infty \exp(-\theta x) f(x)\, dx < \infty, \quad \textit{for any } \theta > 0, \tag{3.83}$$

then one can show that if one has $\rho \in [0, 1)$ *then the Laplace transform of the loss random variable* $X \sim F(x)$ *satisfies the following asymptotic equivalence*

$$\Psi(\theta) \sim \frac{\Gamma(1-\rho)}{\theta} f\left(\frac{1}{\theta}\right), \quad as \quad \theta \downarrow 0. \tag{3.84}$$

This can be rewritten as the density function for the loss random variable is asymptotically equivalent to a scaling of the Laplace transform according to

$$f(x) \sim \frac{\Psi\left(1/x\right)}{x\,\Gamma(1-\rho)}, \quad as \quad x \to \infty. \tag{3.85}$$

In the following chapters, it will also be useful to recall the following property of regular variation of a density function given in Omey and Willekens (1986) and Borovkov and Borovkov (2010) relating the severity distribution tail, severity density and the index of regular variation.

Lemma 3.5 *If a severity distribution function* $f \in RV_{-\alpha}$ *for some* $\alpha \in (1, 2)$*, then the following asymptotic result is satisfied*

$$\lim_{x\to\infty} \frac{f_X(x) \int_0^x (1 - F_X(y))\,dy}{(1 - F_X(x))^2} \to \frac{\alpha - 1}{2 - \alpha}. \tag{3.86}$$

3.4 Alternative Classifications of Heavy-Tailed Models and Tail Variation

It will be relevant in future sections to also consider the notion of consistently varying tails or intermediate regular variation for a distribution F which will have a class denoted by $F \in \mathscr{C}$ and given according to Definition 3.20, see discussion in Cline (1994) and Robert and Segers (2008).

Definition 3.20 (Consistent Variation) *A distribution function* F *is of consistent variation* $(F \in \mathscr{C})$ *iff*

$$\lim_{\gamma\uparrow 1} \text{limsup}_{x\to\infty} \frac{\overline{F}(x\gamma)}{\overline{F}(x)} = 1. \tag{3.87}$$

Furthermore, one can show that the class of consistently varying distribution functions will contain, as a sub-class, the class of regularly varying distribution functions, see Cline and Samorodnitsky (1994). One can also consider a related sub-classes of regularly varying severity models known as the *class of smoothly varying functions* given in Definition 3.21, see Tarov (2004) and Barbe and McCormick (2009a, p. 6).

Definition 3.21 (Smoothly Varying Function) *A real measurable function* $f \in SR_{-\rho}(m)$ *is smoothly varying with index* $-\rho$ *and order* m *if it is eventually (asymptotically)* m*-times continuously differentiable and the* m*-th derivative* $D^m f(x) = f^{(m)}(x)$ *is regularly varying with*

index $f^{(m)} \in RV_{-\rho-m}$. Furthermore, for any non-integer value $u > 0$, with $u = m + r$ for $m \in \mathbb{N}$ and $r \in [0, 1)$, then a function $f(x)$ is smoothly varying with index $-\rho$ and order u if $f \in SR_{-\rho,m}$ and the following limit holds

$$\lim_{\delta \to 0} \text{limsup}_{x \to \infty} \sup_{0<|\lambda|<\delta} \frac{f^{(m)}(x(1-\lambda)) - f^{(m)}(x)}{|\lambda|^r f^{(m)}(x)} = 0. \tag{3.88}$$

One can also state the smooth variation theorem which provides asymptotic bounds on the severity density, see Bingham *et al.* (1989).

Theorem 3.8 (Smooth Variation Theorem) *If a severity density model has a regularly varying right tail $f \in RV_\rho$, then there exist functions f_1 and f_2 with $f_2 \in SR_\rho$ and asymptotic equivalence $f_1 \sim f_2$ such that in some neighbourhood of infinity, the following bounds apply $f_1 \leq f \leq f_2$.*

Corollary 3.4 *If the severity density has a regularly varying right tail, $f \in RV_\rho$, then there exists a function $g \in SR_\rho$ with $g \sim f$.*

This further characterization of a sub-class of regularly varying severity models is highly relevant as it allows one to make some comments on integrals of tail functionals with respect to such severity distributions. In particular, one can show the following is true for two smoothly varying functions when utilized to construct a product function, see Bingham *et al.* (1989, p. 47).

Lemma 3.6 (Products of Smoothly Varying Functions) *Given two smoothly varying functions $f \in SR_\rho$ and $\phi \in SR_\alpha$, the product of these two functions is also smoothly varying such that $f(x)\phi(x) \in SR_{\rho+\alpha}$.*

Remark 3.13 (Implications for Capital Approximations) *This product closure for the family of smoothly varying functions, when combined with convolution closure properties of regularly varying functions is particularly useful when integrating tail functionals of compound processes, such as would be required when calculating capital approximations under certain risk measures such as Expected shortfall or Spectral Risk Measures.*

Building on these sub-classes of regularly varying functions, one can also define a broader class of functions for heavy-tailed loss models under the notion of tail variation such as dominantly varying, subversively varying and long-tailed severity models, see discussion in Bardoutsos and Konstantinides (2011).

Definition 3.22 (Dominantly Varying Tail) *A severity distribution function F is said to have a dominantly varying tail if it satisfies the asymptotic condition that*

$$\text{limsup}_{x \to \infty} \frac{\overline{F}(ux)}{\overline{F}(x)} < \infty, \tag{3.89}$$

for any $u \in (0, 1)$. An alternative equivalent relationship is to consider

$$\text{liminf}_{x \to \infty} \frac{\overline{F}(ux)}{\overline{F}(x)} > 0, \tag{3.90}$$

for any $u > 1$.

This notion of dominated variation is interesting to consider for the following reasons discussed in Goldie (1978). It is well known that for a severity model with a positive support, if the distribution F has a tail $\overline{F}$ which is regularly varying, then in this case, it will imply that F is in the family of sub-exponential distributions. Alternatively, if the tail of the severity distribution $\overline{F}$ is instead merely of dominated variation, then it is no longer the case that the severity distribution F needs to be in the family of sub-exponential models. Borovkov and Borovkov (2002) refer to the class of dominantly varying functions by the term upper-power functions. They state that a distribution has dominantly varying tails, that is, an upper-power function tail, if the loss distribution can be shown to have a tail which is asymptotically locally constant (long tailed) and for any $\gamma > 1$, there exists a constant $c(\gamma) > 0$ such that

$$F(\gamma x) > c_{\gamma} F(x), \quad x > 0, \tag{3.91}$$

with c_{γ} is bounded away from zero on any interval $(1, \gamma)$ with $\gamma < \infty$.

The following properties of distribution functions with dominantly varying tail functions are known, see Proposition 3.22 and additional results in Borovkov and Borovkov (2002).

Definition 3.23 (Properties of Loss Distributions with Dominantly Varying Tails) *If a distribution function has a dominantly varying right tail function, that is, if $\overline{F}$ is dominantly varying, then the distribution $F \in \mathcal{F}$ is also sub-exponential.*

Definition 3.24 (Subversively Varying Tail) *A severity distribution function F is said to have a subversively varying tail if it satisfies the asymptotic condition that*

$$\text{limsup}_{x \to \infty} \frac{\overline{F}(ux)}{\overline{F}(x)} < 1, \tag{3.92}$$

for any $u > 1$.

Previously, when introducing a characterizing feature used to study the sub-exponential family of loss distributions, the concept of asymptotically locally constant distributions was defined, see Definition 3.13. It turns out that this feature is also known to characterize a family of heavy-tailed distributions which have right tails, which are asymptotically locally constant; in such cases, one obtains the class of long-tailed distributions.

Definition 3.25 (Long-Tailed Distributions) *A severity distribution function $\overline{F}$ is said to be long tailed if it satisfies the asymptotic condition that*

$$\lim_{x \to \infty} \frac{\overline{F}(x - y)}{\overline{F}(x)} = 1, \tag{3.93}$$

for all constants $y \in \mathbb{R}$. Equivalently, one can state this by saying that $\overline{F}(x - y)$ is asymptotically equivalent to $\overline{F}(x)$, that is, $\overline{F}(x - y) \sim \overline{F}(x)$.

Remark 3.14 *Long-tailed distributions are heavy tailed, and furthermore, the condition of a severity distribution being long tailed implies a degree of smoothness in the tail function of $\overline{F}$ which may not hold for all heavy-tailed severity models that are not long tailed.*

One further implication of long-tailed distribution functions relates to the asymptotic behaviour of tail convolutions for such distributions, obtained when calculating tail probabilities for sums of independent random variables. A particularly useful result for long-tailed loss models is given in Lemma 3.1, see for instance Foss *et al.* (2013).

Lemma 3.7 (Asymptotics of Tail Convolutions for Long-Tailed Distributions) *Consider two independent loss random variables $X_1 \sim F_1$ and $X_2 \sim F_2$, with F_1 in the class of long-tailed distributions. Furthermore, assume there is a function $h(x)$ for which F_1 is considered flat or insensitive to, that is, the distribution F_1 satisfies the h-insensitive condition, which means that as $x \to \infty$, the condition*

$$\sup_{|y| \leq h(x)} |F_1(x+y) - F_1(x)| = o(F_1(x)) \tag{3.94}$$

holds uniformly in $|y| \leq h(x)$. Then the following tail convolution asymptotic applies for any distribution F_2,

$$\begin{aligned} &\int_{-\infty}^{h(x)} \overline{F}_1(x-y) dF_2(y) \sim \overline{F}_1(x), \\ &\int_{x-h(x)}^{\infty} \overline{F}_2(x-y) dF_1(y) \sim \overline{F}_1(x). \end{aligned} \tag{3.95}$$

Remark 3.15 *If one considers a long-tailed distribution F which is h-flat or h-insensitive to a function $h(x) = \lambda x$ for $\lambda > 0$, then such distributions F are also those which have slowly varying tails at infinity, that is, ultra-heavy-tailed loss distributions.*

Furthermore, one can also verify that a loss distribution is long tailed through the result in Lemma 3.8, see Borovkov and Borovkov (2002).

Lemma 3.8 (Long-Tailed Distribution Verification via a Lower Bound) *Consider two loss distributions F and G with G non-degenerate at the origin and the following distributional lower bound for the tails $\overline{G} \leq c\overline{F}$ for some $c < \infty$. Furthermore, if F satisfies the condition that*

$$\text{limsup}_{x \to \infty} \frac{\overline{F \star G}(x)}{\overline{F}(x) + \overline{G}(x)} \leq 1, \tag{3.96}$$

then F is long tailed.

In addition to long-tailed distributions, one can also talk about long-tailed functions, such as a density function, f as given in Definition 3.26.

Definition 3.26 (Long-Tailed Functions) *A function f is long tailed iff for $x > 0$ one has function $f(\ln(x))$ being slowly varying at infinity, that is, for any constant λ, one has that*

$$\lim_{x \to \infty} \frac{f(\ln \lambda + \ln x)}{f(\ln x)} = 1. \tag{3.97}$$

To verify that a particular loss density is a long-tailed function, one can consider its 'sensitivity' asymptotically to a strictly positive non-decreasing function as specified in Proposition 3.4, see discussions in Foss *et al.* (2013).

Proposition 3.4 (Long-Tailed Function Verification) *If the function f is long tailed, then there exists a strictly positive, non-decreasing function h satisfying that $h(x) \to \infty$ as $x \to \infty$, such that one has the long-tailed function satisfying the h-insensitive condition, where as $x \to \infty$ the condition*

$$\sup_{|y| \leq h(x)} |f(x+y) - f(x)| = o(f(x)), \tag{3.98}$$

holds uniformly in $|y| \leq h(x)$.

The long-tailed severity model case is particularly interesting as members of this family have the property that the distribution of a random variable $X \sim F$ is said to have a long right tail if for all $\lambda > 0$,

$$\lim_{x \to \infty} \mathbb{P}\mathrm{r}[X > x + \lambda | X > x] = 1, \tag{3.99}$$

or equivalently put in terms of asymptotic order it means that

$$\overline{F}(x + \lambda) \sim \overline{F}(x). \tag{3.100}$$

Therefore, the interpretation of a right long-tailed distributed quantity is that if the long-tailed quantity exceeds some high level, the probability approaches 1 that it will exceed any other higher level. Put simply if you know the loss amounts are significant then the realized losses from such a severity model are probably worse than you think.

Remark 3.16 *One can show that all long-tailed distributions are heavy tailed, but the converse is false. In addition, one has that all sub-exponential distributions are long tailed, but examples can be constructed of long-tailed distributions that are not sub-exponential.*

One also has the following properties of long-tailed distributions given in Lemma 3.9, see discussions in Foss *et al.* (2013, lemma 2.16).

Lemma 3.9 (Properties of Long-Tailed Functions and Distributions) *Consider long-tailed functions $f_1, \ldots, f_n$, then the following conditions hold:*

- *Given constants c_1 and $c_2 > 0$, the function $f_1(c_1 + c_2 x)$ will remain in the class of long-tailed functions.*
- *The aggregate of linearly weighted long-tailed functions will remain long tailed. That is, given positive constants $c_i > 0$ for all $i \in \{1, 2, \ldots, n\}$, one has*

$$f \sim \sum_{i=1}^{n} c_i f_i, \tag{3.101}$$

  also being a long-tailed function.
- *The product of long-tailed functions is also a long-tailed function.*
- *The minimum of long-tailed functions is also a long-tailed function.*
- *The maximum of long-tailed functions is also a long-tailed function.*
- *A distribution function F is long tailed iff the tail function of the distribution $\overline{F}$ is also long tailed;*

- *If a distribution F is long tailed, then as $x \to \infty$, for any $y > 0$, one has*

$$F((x, x+y]) = o(\overline{F}(x)); \tag{3.102}$$

- *Consider the long-tailed distribution F with hazard function $R(x)$, then as $x \to \infty$ one has $R(x+1) - R(x) \to 0$.*

Having defined these different families of tail behaviour in a severity distribution or density, we note the following relationships between these families, see discussion in Geluk and Tang (2009).

Definition 3.27 (Properties of Asymptotically Locally Constant Distributions) *The following properties of asymptotically locally constant distributions can be shown:*

1. *A distribution with an asymptotically locally constant tail (long-tailed distribution) will have a convergence*

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1, \tag{3.103}$$

   for all constants $y \in \mathbb{R}$ that is uniform in y on any bounded fixed interval.
2. *A long-tailed distribution is a distribution with an asymptotically locally constant right tail iff eventually for some $x > x_0$ the tail of the loss distribution can be represented by*

$$\overline{F}(x) = c(x) \exp\left(\int_{x_0}^{x} \varepsilon(u)\, du \right), \quad x \geq x_0, \tag{3.104}$$

   where the functions $c(x)$ and $\varepsilon(x)$ are measurable and $c(x) \to c \in (0, \infty)$ and $\varepsilon(x) \to 0$ as $x \to \infty$.
3. *If two loss distributions are each long tailed, then $\overline{F}_1(x) + \overline{F}_2(x)$ is also a long-tailed distribution, $\overline{F}_1(x)\overline{F}_2(x)$ will be a long-tailed distribution.*
4. *If $F(x)$ is a long-tailed distribution, then its asymptotically locally constant right tail can be represented as $x \to \infty$ by*

$$\overline{F}(x) = \exp(-l(x)), \quad l(x) = o(x) \ \ as \ \ x \to \infty. \tag{3.105}$$

One can also show that asymptotically locally constant distributions, such as the long-tailed class of heavy-tailed distributions, satisfy the following properties given in Proposition 3.26, see Borovkov and Borovkov (2002, theorem 1.2.4).

Remark 3.17 (Relating the Families of Severity Models by Tail Behaviour) *The following relationships between the different families of severity distributions, classified by their right tail behaviour, hold:*

1. *Firstly, the class of sub-exponential distributions is larger than the class of regularly varying distributions and one can observe the relationship through the result in Lemma 3.8, see Jessen and Mikosch (2006, lemma 3.2).*

2. *In addition the class of sub-exponential distributions is contained in the class of asymptotically locally constant functions. Secondly, it is well known that the class of smoothly varying functions, functions with continuous derivatives being regularly varying at infinity, is a sub-class of regularly varying functions.*
3. *The intersection between the family of dominantly varying, subversively varying and sub-exponential-tailed distributions is contained in the family comprised the intersection between the dominantly varying tailed functions and the long-tailed functions. Furthermore, these sub-families formed from the intersections are themselves contained in the family of sub-exponential models which is itself contained in the family of long-tailed distributions.*
4. *One has the following sub-functional families relationships between the Regularly Varying (RV) functions, the extended regular variation functions (ER) and the O-regularly varying functions (OR) given by* $RV \subset ER \subset OR$.
5. *The class of dominantly varying distributions* $\mathscr{D}$ *is larger than the class of consistently varying distributions* $\mathscr{C}$ *with the intersection between the long-tailed distributions* $\mathscr{L}$ *and the Dominantly varying distributions containing the class of Consistently varying distributions* $\mathscr{C} \subset \mathscr{L} \cap \mathscr{D}$.
6. *If the distribution satisfies* F *that it is sub-exponential* $F \in \mathscr{F}$, *then the distribution tail satisfies membership of the class of long-tailed distributions* $\overline{F} \in \mathscr{L}$. *Note, in general, the converse of this statement will not hold.*
7. *If* $\overline{F}$ *is in the class intersection of O-regularly varying functions and long-tailed functions* $\overline{F} \in OR \cap \mathscr{L}$, *then this implies that the severity distribution is sub-exponential* $F \in \mathscr{F}$.
8. *Summarizing the relationships, one has the following functional family relationships* $RV \subset \mathscr{D} \cap \mathscr{L} \subset \mathscr{F} \subset \mathscr{L}$.

3.5 Extended Regular Variation and Matuszewska Indices for Heavy-Tailed Models

Using this definition for the class of regularly varying functions, the following extended classes known as the *extended regularly varying functions* (ER) and *O-regularly varying functions* (OR) are given in Definition 3.28. First, we define the following properties for the severity density, using the notation Bingham *et al.* (1989), the lim sup and lim inf according to Equations 3.106 and 3.107 for $\lambda > 0$

$$f^*(\lambda) = \text{limsup}_{x\to\infty} \frac{f(\lambda x)}{f(x)} \tag{3.106}$$

and

$$f_*(\lambda) = \text{liminf}_{x\to\infty} \frac{f(\lambda x)}{f(x)} \tag{3.107}$$

with the relationship that $f_*(\lambda) = 1/f^*(1/\lambda)$.

Definition 3.28 (Extended and O-Type Regular Variation) *The class of extended regularly varying functions is the set of positive measurable functions* $f \in ER$ *satisfying for some constants* c, d *the relationship*

$$\lambda^d \leq f_*(\lambda) \leq f^*(\lambda) \leq \lambda^c, \ \forall \lambda \geq 1. \tag{3.108}$$

The class of O-regularly varying functions is the set of positive measure functions $f \in OR$ satisfying

$$0 < f_*(\lambda) \leq f^*(\lambda) < \infty, \ \forall \lambda \geq 1. \tag{3.109}$$

Another way to define a distribution function as O-regularly varying is given in Definition 3.29, see Geluk (1992).

Definition 3.29 (O-Regular Variation of Distribution Functions) *A distribution function F that is concentrated on $[0, \infty)$ is of O-Regular variation $F \in OR(\alpha)$ if there exists a positive function α which is O-regularly varying, that is, measurable, eventually positive and satisfying for $x > 0$ the condition*

$$\lim_{t \to \infty} \frac{\alpha(tx)}{\alpha(t)} < \infty, \tag{3.110}$$

where $\alpha(x) = o(\overline{F}(x))$ as $x \to \infty$ and

$$\text{limsup}_{x \to \infty} \frac{\overline{F}\left(\frac{1}{2}x\right) - \overline{F}(x)}{\alpha(x)} < \infty. \tag{3.111}$$

Analogously, to the O-regular variation, it is also often useful to define the notion O-sub-exponentiality, which is given by Definition 3.30.

Definition 3.30 *A distribution function F that is concentrated on $[0, \infty)$ is of O-Sub-exponentiality $F \in OS(\alpha)$ if there exists a positive function α which is O-regularly varying, that is, measurable, eventually positive and satisfying, for $x > 0$, the condition*

$$\lim_{t \to \infty} \frac{\alpha(tx)}{\alpha(t)} < \infty, \tag{3.112}$$

where $\alpha(x) = o(\overline{F}(x))$ as $x \to \infty$ and

$$\text{limsup}_{x \to \infty} \frac{\overline{F^{(2)*}}(x) - 2\overline{F}(x)}{\alpha(x)} < \infty. \tag{3.113}$$

Using these definitions, one can show the following bounds for the partial sum of n-losses each of which is i.i.d. from a severity distribution $F \in OR(\alpha)$ given in Proposition 3.5, see Geluk (1992).

Proposition 3.5 (Partial Sum Bounds for O-Regularly Varying Severities) *If a distribution F satisfies that it is a member of the O-regularly varying class $F \in OR(\alpha)$, then there exists constants δ_n for $n \geq 2$ such that the following bound on the tail of the distribution of the partial sum in terms of the tail of the severity distribution is obtained*

$$\overline{F^{(n)*}}(x) - n\overline{F}(x) \leq \delta_n \alpha(x), \ \ x > 0. \tag{3.114}$$

Using this result, Geluk (1992) shows the following asymptotic big-Oh order of approximation for the n-fold convolution approximation of the partial sum distribution tail by the n-th power of the severity distribution tail given in Proposition 3.6.

Proposition 3.6 (Partial Sum Tail Distribution Approximation via Severity Tail Distribution Powers) *Consider the distribution function for the loss random variables $F(x)$ that satisfies the condition*

$$\text{limsup}_{x\to\infty} \frac{\overline{F}\left(\frac{1}{2}x\right) - \overline{F}(x)}{\overline{F}(x)^3} < \infty, \tag{3.115}$$

then one has the order of approximation

$$\overline{F^{(n)*}}(x) - \overline{F^n}(x) = O\left(\overline{F}(x)^3\right). \tag{3.116}$$

Geluk (1992) also provides an analogue of the famous convolution root closure condition for sub-exponential distributions for the case of O-Regularly varying distributions as given in Proposition 3.7.

Proposition 3.7 (Convolution Root Closure of O-Regular Variation Distributions) *Consider a severity distribution F that satisfies that its n-fold convolution with itself $F^{(n)*}$ is of O-regular variation $F^{(n)*} \in OR(\alpha)$ for some $n \geq 2$ and the tail of the n-fold convolution satisfies $F^{(n)*}(x)^2 = O(\alpha(x))$ as $x \to \infty$ then one has that the severity distribution is also of O-regular variation $F \in OR(\alpha)$. Hence, the severity distribution F is closed under convolution.*

In addition, the results on O-regular variation and extended regular variation can also be utilized to state the definition of the Matuszewska index provided in Definition 3.31, see Matuszewska (1964) and Bingham *et al.* (1989, p. 68).

Definition 3.31 (Matuszewska Index) *Let f be a positive density function, then the upper Matuszewska index, denoted by $\alpha(f)$, is given as $x \to \infty$ by the infimum of the α values such that there exists a constant $C = C(\alpha)$ where for each $\Lambda > 1$ one has*

$$\frac{f(\lambda x)}{f(x)} \leq C\{1 + o(1)\}\lambda^{\alpha} \quad \textit{uniformly in } \lambda \in [1, \Lambda]. \tag{3.117}$$

The lower Matuszewska index, denoted by $\beta(f)$, is analogously given as $x \to \infty$ by the supremum of β values for which for some constant $D > 0$ and for all $\Lambda > 1$ one has

$$\frac{f(\lambda x)}{f(x)} \geq D\{1 + o(1)\}\lambda^{\beta} \quad \textit{uniformly in } \lambda \in [1, \Lambda]. \tag{3.118}$$

The following relationship between Matuszewske index is known for positive functions f

$$\beta(f) = -\alpha\left(\frac{1}{f}\right). \tag{3.119}$$

One can show the following properties of the Matuszewska Index for a positive function f given in Lemma 3.10, see Bingham *et al.* (1989, p. 71).

Lemma 3.10 *Consider a positive function f, then the following properties w.r.t. the Matuszewska index's sign can be shown:*

1. *f has bounded increase $f \in BI$ if $\alpha(f) < \infty$;*
2. *f has bounded decrease $f \in BD$ if $\beta(f) > -\infty$;*
3. *f has positive increase $f \in PI$ if $\beta(f) > 0$;*
4. *f has positive decrease $f \in PD$ if $\alpha(f) < 0$.*

Of direct relevance to the results developed in future chapters on higher order asymptotic tail expansions will be the extension of the concept of Matuszewska indices which was further developed for distribution function tails in Cline and Samorodnitsky (1994) who provided the statements in Lemma 3.11.

Lemma 3.11 (Matuszewska Indices for Distribution Functions) *Given a severity distribution F, the upper Matuszewska index for the tail of the distribution ($\overline{F}(x) = 1 - F(x)$) denoted by $\gamma_{\overline{F}}$ is given by*

$$\gamma_{\overline{F}} = \inf \left\{ -\frac{\ln \overline{F}_*(u)}{\ln u} : u > 1 \right\} = -\lim_{u \to \infty} \frac{\ln \overline{F}_*(u)}{\ln u}. \tag{3.120}$$

The lower Matuszewska index is given for the tail of a distribution analogously by

$$\delta_{\overline{F}} = \sup \left\{ -\frac{\ln \overline{F}^*(u)}{\ln u} : u > 1 \right\} = -\lim_{u \to \infty} \frac{\ln \overline{F}^*(u)}{\ln u}. \tag{3.121}$$

Analogous definitions can also be developed for severity density functions.

When the upper and lower Matuszewska indices are finite for the tails of a distribution function one may state the following bounds in Proposition 3.8.

Proposition 3.8 *Given a severity distribution function F with a finite upper Matuszewska index $\gamma_{\overline{F}} < \infty$, there exist constants C_1 and x_0 such that the bound*

$$\frac{\overline{F}(x)}{\overline{F}(y)} \leq C_1 \left(\frac{x}{y} \right)^{-\gamma} \tag{3.122}$$

holds for all $x \geq y \geq x_0$ and $\gamma_{\overline{F}} < \gamma < \infty$. Furthermore, if the lower Matuszewska index is finitely positive $\delta_{\overline{F}} > 0$, then there exist constants C_2 and x_0 such that

$$\frac{\overline{F}(x)}{\overline{F}(y)} \geq C_2 \left(\frac{x}{y} \right)^{-\delta} \tag{3.123}$$

holds for all $x \geq y \geq x_0$ and $0 < \delta < \delta_{\overline{F}^*}$.

CHAPTER FOUR

Flexible Heavy-Tailed Severity Models: α-Stable Family

In this chapter, we present advanced discussion for risk modellers relating to flexible distributions. The intention of this chapter is to provide a clear understanding of the key features of a very flexible and widely applicable class of models in the operational risk (OpRisk) and insurance context, provided by the α-stable family.

In order to develop a basic understanding of this important family of severity models, we first present several fundamental properties of the primary building block of all these models, the α-stable family of models. The family of univariate α-stable severity models describes not only heavy tails and asymmetric behaviour but also the dependency on four parameters that allows for a wide range of skewness and kurtosis features in loss data to be modeled appropriately. In addition, as will be discussed later, the domain of attraction for such families of models will naturally correspond to considering the distributions from which limits of sums of i.i.d. random variables will converge weakly (in distribution) to laws which are non-degenerate.

To proceed, we first present a note on some of the notations utilized in this chapter:

- The α-stable severity distribution, denoted by $S_\alpha(x;\beta,\gamma,\delta;\text{TYPE})$, is parameterized by parameters (generically denoted) α, β, γ and δ. The different parameterizations of this family of distributions will be made explicit with a subscript index and also by the specification of a label for TYPE.
- There are many choices for parameters and representations in the α-stable family, we will take extra caution below to avoid any 'comedy of errors' in presenting the results, note, initially when deriving the A-Type representation of Zolotarev, we will work with the original parameterization using parameters α, β, λ and γ, and then we will convert to show the relationship between all popular parameterizations and use a universal notation of α, β, γ and δ with subscripts to indicate the parameterizations type.
- The characteristic function for a random variable $X \sim F_X(x)$ will be denoted by $\Phi_X(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta X)]$, where $\mathrm{i} = \sqrt{(-1)}$ is a unit imaginary number.

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

- When talking about sequences of random variables and weak convergence (convergence in distribution) of the distributions in such a sequence to a family of distributions such as the α-stable family, we will refer to the concept of domain of attraction. The domain of attraction for such a sequence of random variables is the subset of all such distributions for these random variables that result in the weak convergence in distribution to a limiting distribution, denoted by $G(x)$, or family of limiting distributions (when it exists). The domain of all such distributions that produce this limiting convergence is denoted by the domain of attraction notation $\mathscr{D}(G)$.
  - In the particular case of the α-stable domain of attraction, we will refer to this particular case of weak convergence of appropriately scaled and translated sums of random variables by the domain of attraction for the distributions that will have an α-stable limit by the special notation $\mho$. In addition, when we say $\mathscr{D}(G) \subset \mho$, we refer to domains of attraction for sub-family members of the α-stable family such as the case $\alpha = 2$; this would be the standard Gaussian domain of attraction in the central limit theorem, that is, the set of all distributions in this domain correspond to those for which linear combinations of the random variables of these distributions converge weakly to a Gaussian, after appropriate scaling of the summands by a deterministic constant sequence.
- The L-class of distributions (defined later) will be denoted by the bold-faced $\mathbb{L}$ notation.
- When discussing Lévy–Khintchine representations for the characteristic function of a random variable, the Lévy triplet will be characterized by a Lévy measure denoted generically by $W(dx)$, the drift will generically be denoted by a and the volatility component is generically denoted by σ^2. In general, any exceptions to this notation for the Lévy triple will be made clear in the text.

4.1 Infinitely Divisible and Self-Decomposable Loss Random Variables

When introducing the family of heavy-tailed severity models corresponding to the flexible four-parameter α-stable distributions, it will first be instructive to recall some fundamental definitions that motivate many of the properties of the stable family. These include the following concepts:

- Properties of characteristic functions;
- n-divisibility and infinite divisibility of random variables;
- Self-decomposability;
- Domains of attraction and limit theorems for triangular arrays.

4.1.1 BASIC PROPERTIES OF CHARACTERISTIC FUNCTIONS

We start by recalling the basic definition of a characteristic function and key properties in Definition 4.1.

Definition 4.1 (Characteristic Function) *The characteristic function for a scalar loss random variable $X \sim F(x)$ is a mapping $\Phi_X(\theta) : \mathbb{R} \to \mathbb{C}$ given by one of the following representations*

$$\begin{aligned}\Phi_X(\theta) &:= \mathbb{E}[\exp(\mathrm{i}\theta X)] \\ &= \int_{-\infty}^{\infty} \exp(\mathrm{i}\theta x)\, dF(x) \\ &= \int_{-\infty}^{\infty} \cos(\theta x)\, dF(x) + \mathrm{i}\int_{-\infty}^{\infty} \sin(\theta x)\, dF(x) \\ &= \int_{-\infty}^{\infty} \exp(\mathrm{i}\theta x) f(x)\ dx \\ &= \int_{0}^{1} \exp(\mathrm{i}\theta F^{-1}(p))\ dp\end{aligned} \tag{4.1}$$

depending if one has access to the knowledge of the distribution function $F(x)$, the density function $f(x)$ or the quantile function $F^{-1}(p)$. In addition, it is apparent that the characteristic function is clearly a version of the (inverse) Fourier transform (FT) when the probability density function (pdf) $f(x)$ exists.

An important property of characteristic functions that will be of relevance to results considered in this chapter for univariate α-stable loss random variables will be the Lévy–Cramer continuity theorem given in Theorem 4.1.

Theorem 4.1 (Lévy–Cramer Continuity Theorem) *Consider loss random variables $X_1, X_2, \ldots, X_n$ for $n = 1, 2, \ldots$ with characteristic functions $\Phi_{X_i}(\theta)$, then the following holds:*

- *If a loss random variable X_n converges in distribution to another loss random variable X, that is, $X_n \xrightarrow{d} X$ as $n \to \infty$, then one has the following limit of the characteristic function sequence pointwise where $\lim_{n\to\infty}\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ for $\theta \in \mathbb{R}$.*
- *Conversely, if the characteristic function limit $\lim_{n\to\infty}\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ for $\theta \in \mathbb{R}$, where $\Phi_X(\theta)$ is a function that is continuous at zero, then $\Phi_X(\theta)$ is the characteristic function of a random variable X and $X_n \xrightarrow{d} X$ as $n \to \infty$.*

It is also well known that Polya's criterion provides the sufficient condition for a function mapping $\mathbb{R} \mapsto \mathbb{C}$ to be a characteristic function for a random variable, see Theorem 4.2, see discussion in, for instance, Lukacs (1970).

Theorem 4.2 (Polya's Sufficient Condition for a Characteristic Function) *Consider a function $\Phi : \mathbb{R} \mapsto \mathbb{R}_+$, which is even and continuous and satisfies that $\Phi(0) = 1$, then the following holds:*

1. *If the function Φ is non-decreasing and convex on $(0, \infty)$, then Φ is the characteristic function of a symmetric distribution on $\mathbb{R}$.*
2. *If Φ is non-increasing and convex on $(0, \pi)$ and 2π-periodic, then Φ is the characteristic function of a symmetric distribution.*

Furthermore, throughout this chapter, it will be useful to consider the behaviour of the logarithm of the characteristic function which has the following property in Lemma 4.1.

Lemma 4.1 (Logarithm of a Characteristic Function) *A characteristic function $\Phi_X(\theta)$ for a loss random variable X which is non-vanishing anywhere on the real line $\mathbb{R}$ has a well-defined logarithm given by*

$$\ln\Phi_X(\theta) = \ln|\Phi_X(\theta)| + \mathrm{i}\ \arg\{\Phi_X(\theta)\}, \quad \theta \in \mathbb{R}, \tag{4.2}$$

with an interval say $[-\pi, \pi)$ for the values of the phase argument $\arg\{\Phi_X(\theta)\}$.

Hence, using this result, one can define for instance that the power can be defined by $\Phi_X(\theta)^t := \exp(t \ln\Phi_X(\theta))$.

Another result of direct relevance to the representation of the α-stable loss models characteristic function is discussed in Steutel and Van Harn (2003) and detailed in Theorem 4.3.

Theorem 4.3 (Poisson Summation and Characteristic Functions) *Consider a characteristic function $\Phi_X(\theta)$ for a random variable X, which is non-negative, integrable and non-increasing on $(0, \infty)$ with a corresponding continuous density $f_X(x)$. Then the following relationship holds:*

$$\sum_{k\in\mathbb{Z}} \Phi_X(\theta + 2k\pi) = \sum_{k\in\mathbb{Z}} f_X(k)\exp(\mathrm{i}\theta k), \quad \theta \in \mathbb{R}. \tag{4.3}$$

The following is a non-complete set of properties, which will be useful to recall in the following discussions, that the characteristic function will satisfy:

1. The characteristic function of any loss random variable always exists. The reason for this is that it corresponds to the integral of a bounded continuous function over a measurable space;
2. A characteristic function is uniformly continuous on the entire space over which it is defined;
3. A characteristic function is bounded and non-vanishing in a region around zero, with $\Phi_X(0) = 1$;
4. A characteristic function is Hermitian.
5. A characteristic function of a symmetric random variable is a strictly real-valued and even function, where a real-valued function $f(x)$ is even if the following equation holds for all x and $-x$ in its domain: $f(x) = f(-x)$.
6. A characteristic function is non-negative definite such that for any integer $n \in \mathbb{J}^+$ and any real numbers $\theta_1, \ldots, \theta_n$ and complex numbers $\lambda_1, \ldots, \lambda_n$, one has the condition

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \Phi_X(\theta_i - \theta_j)\lambda_i\overline{\lambda}_j \geq 0. \tag{4.4}$$

7. If one has n random variables $X_1, X_2, \ldots, X_n$, which are all independent, then the characteristic function of the sum of these random variables has the property

$$\Phi_{\sum_{i=1}^n X_i}(\theta) = \prod_{i=1}^{n} \Phi_{X_i}(\theta) \tag{4.5}$$

for any $n = 1, 2, \ldots$. That is, characteristic functions transform convolutions for distributions and densities of aggregate losses into products of the severity distribution of the loss random variable.

8. The characteristic function of $-X$ will correspond to the complex conjugate $\overline{\Phi}_X(\theta)$ of the characteristic function $\Phi_X(\theta)$.

9. The random variable X will then have a symmetric distribution if $\Phi_X(\theta)$ is real or, equivalently, $\Phi_X(\theta)$ is an even function.

10. If $\Phi_X(\theta)$ is a characteristic function, then so is $|\Phi_X(\theta)|^2$.

11. Limits of characteristic functions are also characteristic functions, provided these limits are also continuous at the origin.

Note items 2, 3 and 6 are necessary and sufficient for $\Phi_X(\theta)$ to be a characteristic function according to the Bochner–Khinchin theorem, see Billingsley (2008). There is another important theorem relating convergence in distribution of a sequence of random variables as equivalent to pointwise convergence of the corresponding sequence of characteristic functions, known as *Lévy's continuity theorem*; however, this will not be required here, see Williams (1991).

One final property of the characteristic function that is worth recalling is the uniqueness of the (Fourier transform–inverse Pair) characteristic function to represent the probability distribution function, as detailed in Theorem 4.4.

Theorem 4.4 (Uniqueness of the Characteristic and Distribution Function Pair) *For any $a \in \mathbb{R}$ and $b \in \mathbb{R}$ with $\Pr[X = a] = 0$ and $\Pr[X = b] = 0$ and $a < b$, one has*

$$\Pr[a < X \leq b] = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{\exp(-\mathrm{i}\theta a) - \exp(-\mathrm{i}\theta b)}{\mathrm{i}\theta} \Phi_X(\theta)\, d\theta, \tag{4.6}$$

where $\Phi_X(\theta)$ is the characteristic function of the random variable $X \sim F(x)$.

4.1.2 DIVISIBILITY AND SELF-DECOMPOSABILITY OF LOSS RANDOM VARIABLES

A detailed account of infinite divisibility on the real line, including both continuous and discrete-valued random variables can be found in Steutel and Van Harn (2003). Here we summarize a few key definitions and properties that will be of relevance to understanding α-stable loss distributions.

We start this section by recalling the basic definition of decomposability and divisibility. Firstly we note that the *n*-decomposability of a random variable will also imply that it admits lower order i.e. $n-1$, $n-2$ etc. decomposability. However, this property is not true for divisibility of a random variable. In particular, a random variable which is *n*-divisible in general need not be lower order i.e. $n-1$, $n-2$ etc. divisible.

Definition 4.2 (*n*-Decomposable Loss Random Variables) *A random variable X is n-decomposable if there exists independent random variables $\{Y_i\}_{i=1}^{n}$ such that X satisfies*

$$X \stackrel{d}{=} Y_1 + \cdots + Y_n, \tag{4.7}$$

where Y_i's each have their own distribution.

If, in the case, one can say that the random variable in the decomposition are also identically distributed, then one obtains the notion of divisibility.

Definition 4.3 (n-Divisible Loss Random Variables) *A random variable X is n-divisible if there exists independent, identically distributed random variables $\{Y_i\}_{i=1}^n$ such that X satisfies*

$$X \stackrel{d}{=} Y_1 + \cdots + Y_n. \tag{4.8}$$

In the limiting case as $n \to \infty$, one can define the notion of infinite divisibility.

Definition 4.4 (Infinite Divisibility) *A loss random variable X is infinitely divisible if for every $n \in \mathbb{N}$ one can find a sequence of independent, identically distributed random variables $\{Y_{i,n}\}_{i=1}^n$ satisfying*

$$X \stackrel{d}{=} Y_{1,n} + Y_{2,n} + \cdots + Y_{n,n}. \tag{4.9}$$

Infinite divisibility is a property of the distribution function of a random variable. The following properties of infinitely divisible distributions are known:

1. A distribution function F is infinitely divisible iff for every integer $n \in \mathbb{N}$ it is the n-fold convolution of a distribution F_n with itself.
2. A characteristic function $\Phi(\theta)$ is infinitely divisible iff for every integer $n \in \mathbb{N}$ it is the n-fold power of a characteristic function $\Phi_n(\theta)^n$.
3. If a random variable X is infinitely divisible, then scaling such as aX for any $a \in \mathbb{R}$ produces a new infinitely divisible random variable.
4. If two independent random variables X and Y are each infinitely divisible, then the linear combination $X + Y$ is still infinitely divisible.
5. Given a loss random variables distribution $F(x)$, it can always be decomposed according to discrete atom components F_d and a continuous component F_c according to the decomposition result
$$F(x) = wF_d(x) + (1 - w)F_c(x) \tag{4.10}$$
with mixture weight $w \in [0, 1]$. In the context of infinite divisible distributions $F(x)$, one can show that if $F(x)$ admits the above decomposition with $w > 0$, then the discrete component F_d must be an infinitely divisible discretely supported distribution.
6. Any non-degenerate bounded random variable cannot be infinitely divisible.
7. The characteristic function Φ of an infinitely divisible distribution has the property that $\Phi(\theta) \neq 0$ for all $\theta \in \mathbb{R}$ and Φ^p is a characteristic function for all $p > 0$

If one considers the definition of infinite divisibility, it naturally takes the form of an array limit theorem because it states that for infinitely divisible random variables, one can obtain for any $n \in \mathbb{N}$ the representation

$$X \stackrel{d}{=} X_{n,1} + X_{n,2} + \cdots + X_{n,k_n} \tag{4.11}$$

for independent and identically distribution random variables $\{X_{n,j}\}_{j=1}^{k_n}$ as $n \to \infty$ with $k_n \uparrow \infty$. If the assumption of identically distributed random variables in each row of the array is relaxed, so that the random variables $\{X_{n,j}\}_{j=1}^{k_n}$ are only independent, then infinite divisibility is obtained under the additional condition known as uniform almost negligibility of

the sequence of random variables in the sum asymptotically, characterized for each row by the condition

$$\lim_{n\to\infty} \max_{1\le j\le k_n} \Pr(|X_{n,j}| \ge \varepsilon) = 0 \tag{4.12}$$

for all $\varepsilon > 0$.

It is also important to introduce the notion of self-decomposability of a random variable given in Definition 4.5.

Definition 4.5 (Self-Decomposable Random Variables) *A random variable X is called self-decomposable if it can be represented according to*

$$X \stackrel{d}{=} wX + X_w \tag{4.13}$$

for every $w \in [0,1]$ and X independent of random variable X_w whose distribution depends on the value of w.

As will be discussed later in this chapter, a self-decomposable random variable satisfies the following convergence result

$$\lim_{n\to\infty} \frac{1}{B_n}(X_1 + \cdots + X_n - A_n) \xrightarrow{d} X \tag{4.14}$$

for independent random variables $\{X_i\}_{i=1}^n$ for some sequence $A_n \in \mathbb{R}$ and $B_n > 0$.

This leads one to the notion of the domain of attraction, discussed in the Extreme Value Theory in Chapter 2, which we briefly define again in the following section. In this section we discuss the domain of attraction for sums of random variables, as opposed to the definition used in the EVT chapter that considered the maximum of sequences of random variables.

Definition 4.6 (Domain of Attraction of a Random Variable) *Independent and identically distributed (i.i.d) random variables $\{Y_i\}_{i=1}^n$ are said to have a distribution in the domain of attraction of the random variable X's distribution iff they satisfy the weak stability result*

$$\lim_{n\to\infty} \frac{1}{B_n}(Y_1 + \cdots + Y_n - A_n) \xrightarrow{d} X \tag{4.15}$$

for some sequence of $A_n \in \mathbb{R}$ and $B_n > 0$.

To complete this brief introduction to infinite divisibility, it will also be useful to discuss properties of transforms of infinitely divisible distributions, described in Steutel and Van Harn (2003a) as canonical representations. To understand such representations, consider the characteristic function Φ of an infinitely divisible loss random variable; because Φ can be represented by its n-th root $\Phi(\theta) = (\Phi_n(\theta))^n$ for some characteristic function Φ_n, then one has the representation

$$\Phi(\theta) = (\Phi_n(\theta))^n = \exp(n \ln(1 - (1 - \Phi_n(\theta)))). \tag{4.16}$$

From this representation, one can then show that the characteristic function of the infinitely divisible distribution is obtained as the limit

$$\begin{aligned}\Phi(\theta) &= \lim_{n\to\infty} \exp(n(\Phi_n(\theta) - 1)) \\ &= \exp\left(\lim_{n\to\infty} n \int_{\mathbb{R}} (\exp(\mathrm{i}\theta x) - 1)\, dF_n(x)\right).\end{aligned} \tag{4.17}$$

This result is achieved by noting that because an infinitely divisible characteristic function must satisfy that $\Phi(\theta) \neq 0$ but can be represented by the nth root Φ_n for any $n \in \mathbb{N}$, then one must have that $\Phi_n(\theta) \to 1$. In addition, the following asymptotic relationship holds $-\ln(1-z) \sim z$ which gives the required result.

Under this canonical representation, one realizes immediately that the component in the exponent corresponding to $\lim_{n\to\infty}\int_{\mathbb{R}} n dF_n(x)$ will not converge to a bounded function. It is well known that if the characteristic function Φ corresponds to a loss random variable with a distribution which has an atom taking positive mass at the origin given by $p_0 = \mathbb{P}\mathrm{r}[X = 0]$ with no other atoms in the support of $X \sim F(x)$, then the implication is that F_n will have a jump at the origin given by $p_0^{1/n}$. This implies that one can construct a characteristic function for the continuous component of Φ in this single-atom case for random variable X according to

$$\Psi_n(\theta) = \frac{1}{1-p_0^{1/n}}\left(\Phi(\theta) - p_0^{1/n}\right). \tag{4.18}$$

Returning to the challenge of $\lim_{n\to\infty}\int_{\mathbb{R}} n\, dF_n(x)$, one can obtain a well-defined finite limit given by

$$\Phi(\theta) = \exp(\lambda(\lim_{n\to\infty}\Psi_n(\theta) - 1)) = \exp(\lambda(\Psi(\theta) - 1)), \tag{4.19}$$

where this follows by defining $\lambda := -\ln p_0 > 0$ and noting that $\lim_{n\to\infty} n\left(1 - p_0^{1/n}\right) \to \lambda$ and applying the continuity theorem.

The consequence of this canonical representation is profound for the representation of infinitely divisible distributions as it provides the intuition for the following special case result for infinitely divisible characteristic function canonical representations provided in Theorem 4.5 which corresponds to the single atom at the origin case.

Theorem 4.5 *$\Phi : \mathbb{R} \mapsto \mathbb{C}$ is the characteristic function of an infinitely divisible random variable with distribution $X \sim F(x)$, with decomposition given by a single-atom component at the origin F_d and a continuous component F_c*

$$F(x) = wF_d(x) + (1-w)F_c(x) \tag{4.20}$$

with mixture weight $w \in (0,1]$, iff Φ takes a shifted compound Poisson characteristic function form given canonically by

$$\Phi(\theta) = \exp\left[\mathrm{i}\theta\gamma + \lambda\int_{\mathbb{R}}\left(\exp(\mathrm{i}\theta x) - 1\right) dG(x)\right] \tag{4.21}$$

for $\gamma \in \mathbb{R}$, $\lambda > 0$ and distribution G continuous at the origin. (γ, λ, G) is known as the canonical triple which is unique for characteristic function Φ.

When one generalizes to any infinitely divisible distribution with a general decomposition of the distribution function according to

$$F(x) = wF_d(x) + (1-w)F_c(x) \tag{4.22}$$

with $w \in [0,1]$ for any number of atoms in the discrete component, not necessarily at the origin, then one needs to treat the limit $\lim_{n\to\infty}\int_{\mathbb{R}} n\, dF_n(x)$ differently, see detailed

discussions in Lukacs (1970) and the special case presented later for univariate α-stable distributions.

In general, one can develop two possible canonical representations given in Theorems 4.6 and 4.7 depending on the finiteness of the resulting variance of the infinitely divisible random variable, see discussions in Steutel and Van Harn (2003).

Theorem 4.6 (Kolmogorov Canonical Infinitely Divisible Random Variable Characteristic Function) $\Phi : \mathbb{R} \mapsto \mathbb{C}$ *is the characteristic function of an infinitely divisible random variable with distribution* $X \sim F(x)$ *which has a finite non-zero variance, with decomposition given by a general discrete component distribution* F_d *and a continuous component* F_c

$$F(x) = wF_d(x) + (1-w)F_c(x) \tag{4.23}$$

with mixture weight $w \in [0,1]$*, iff* Φ *can be represented in the general form*

$$\Phi(\theta) = \exp\left[\mathrm{i}\theta\mu + \kappa \int_{\mathbb{R}} \frac{1}{x^2} \left(\exp(\mathrm{i}\theta x) - 1 - \mathrm{i}ux\right) dH(x)\right] \tag{4.24}$$

for $\mu \in \mathbb{R}$*,* $\kappa > 0$ *and distribution* H *continuous at the origin.* (μ, κ, H) *is known as the canonical triple, which is unique for characteristic function* Φ*.*

If the variance is not constrained to be finite, one gets the most general form of canonical representation for an infinitely divisible random variable according to Lévy's canonical representation given in the following section.

Theorem 4.7 (Lévy Canonical Infinitely Divisible Random Variable Characteristic Function) $\Phi : \mathbb{R} \mapsto \mathbb{C}$ *is the characteristic function of an infinitely divisible random variable with distribution* $X \sim F(x)$*, with decomposition given by a general discrete component distribution* F_d *and a continuous component* F_c

$$F(x) = wF_d(x) + (1-w)F_c(x) \tag{4.25}$$

with mixture weight $w \in [0,1]$*, iff* Φ *can be represented in the general form*

$$\Phi(\theta) = \exp\left[\mathrm{i}\theta a - \frac{1}{2}\theta^2\sigma^2 + \int_{\mathbb{R}\setminus\{0\}} \left(\exp(\mathrm{i}\theta x) - 1 - \frac{\mathrm{i}ux}{1+x^2}\right) dM(x)\right] \tag{4.26}$$

for $a \in \mathbb{R}$*,* $\sigma > 0$ *and a right continuous non-decreasing function* M *on* $(-\infty, 0)$ *and* $(0, \infty)$ *satisfying* $M(x) \to 0$ *as* $x \to -\infty$ *or* $x \to \infty$ *and satisfies the condition*

$$\int_{(-1,1)\setminus\{0\}} x^2 \, dM(x) < \infty. \tag{4.27}$$

Having introduced some key concepts of infinite divisibility and properties of characteristic functions, we now consider these particularly in the context of heavy-tailed severity models given by the α-stable family of models.

4.2 Characterizing Heavy-Tailed α-Stable Severity Models

The first sub-family of the sub-exponential models we consider is the univariate parametric family of α-stable severity distributions. This sub-family is often associated with heavy-tailed high consequence event modeling in both insurance and risk scenarios, see examples in Feldman and Taqqu (1998) and Peters *et al.* (2011b, 2013). The α-stable family of severity models includes members in which all moments exist and are finite, including the Gaussian distribution through to members such as the Cauchy and Lévy models for which no moments are finite, hence providing possibilities of infinite mean or infinite variance models. In addition, this family of models allows for a diverse range of skewness and asymmetry properties.

The family of α-stable statistical models was largely developed in the 1920s by Paul Lévy; consequently, these models have been further developed in a range of theoretical and applied domains, see Feller (1966) and Zolotarev (1986) for detailed presentations of univariate α-stable models and see Rachev and Mittnik (2000), Samorodnitsky and Taqqu (1997) and Uchaikin and Zolotarev (1999) for recent textbook reviews dealing with their properties and applications. In addition, there have been several well-written summary papers on the properties of α-stable models, such as the work of Janicki and Weron (1994a) which provides a historical perspective on the development of α-stable models as well as a short summary of many relevant properties of such models.

α-Stable distributions have found application in many areas of statistics Peters *et al.* (2010a), finance Peters *et al.* (2011b, 2014) and signal processing engineering as models for impulsive, heavy-tailed noise processes (Mandelbrot, 1960; Fama, 1965; Fama and Roll, 1968; Nikias and Shao, 1995; Godsill, 2000a; Melchiori, 2006). A comprehensive list of such research papers in several fields in which α-stable models are developed is maintained by John Nolan in a bibliography.[1]

The popularity of models constructed with α-stable distributions stems from the properties this family of distributions possesses, which can include infinite variance, skewness and heavy tails (Zolotarev, 1986; Alder *et al.*, 1998; Samorodnitsky and Taqqu, 1994a; Nolan, 2007).

Whilst such models are highly flexible in their skew, kurtosis and tail features, they do pose challenges for working with them in practice. For instance, α-stable distribution families provide no general analytic expressions for the density, median, mode or entropy except for a select few special members. They are uniquely specified by their characteristic function, which has several parameterizations, of which we will discuss a small subset and refer the interested reader to the references of Zolotarev (1986) and Samorodnitsky and Taqqu (1997). α-stable distributions can also be considered as generalizations of the Gaussian distribution and are defined as the class of location-scale distributions which are closed under convolutions, a property that is particularly useful in OpRisk and insurance models, see Peters *et al.* (2011).

We first present some basic concepts that will be utilized throughout this section relating to the definition of an α-stably distributed random variable from first principles. Having defined the fundamental stable properties relating to closure under convolution of this family of distributions, we define a few parameterizations for the characteristic function in the univariate setting applicable to OpRisk and insurance modeling.

We will present in the following sections four equivalent 'characterizations' for the definition of the family of univariate α-stable severity distribution models. The first is sometimes

[1] http://academic2.american.edu/∼jpnolan/stable/StableBibliography.pdf.

referred to as the definition of strictly stable members of the family, see Definition 4.7 from Samorodnitsky and Taqqu (1997).

Definition 4.7 (Characterization 1 of an α-Stable Severity Model) *Consider random variable X and independent copies X_1 and X_2 with the same distribution as X, then random variable X has a stable distribution if for any constants $a_1 \in \mathbb{R}^+$ and $a_2 \in \mathbb{R}^+$, there exists $a_3 \in \mathbb{R}^+$ and $a_4 \in \mathbb{R}$ such that the following equality in distribution is satisfied*

$$a_1 X_1 + a_2 X_2 \stackrel{d}{=} a_3 X + a_4.$$

Remark 4.1 *In situations in which $a_4 = 0$, the random variable X is considered strictly stable and we note that in this situation, the distribution of X is necessarily symmetric.*

In addition, as stated in Samorodnitsky and Taqqu (1997, theorem 1.1.2) the following result, based on Definition 4.7, can be stated for any stable random variables.

Corollary 4.1 *For any stable random variable X, there is a number $\alpha \in (0, 2]$ such that α satisfies*

$$a_3^\alpha = a_1^\alpha + a_2^\alpha.$$

Remark 4.2 *This parameter α will play a very important role in the characterization of a member of the stable family. In particular, it is known as the characteristic exponent of the stable family and dictates the heaviness of the tails of the stable family member. It is directly related to the index of regular variation.*

The second characterization of the α-stable family is an equivalent extension of Characterization 1, for n loss random variables that are independent (not necessarily identically distributed).

Definition 4.8 (Characterization 2 of an α-Stable Severity Model) *Consider random variable X and for a positive integer $n \geq 2$ the independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution as X, then random variable X has a stable distribution if for any constants $c_n \in \mathbb{R}^+$ and $d_n \in \mathbb{R}$, the following equality in distribution is satisfied between the n-fold convolution of the X_1 to X_n copies and the scaling and translation of the original random variables distribution for X according to the relationship*

$$X_1 + X_2 + \cdots + X_n \stackrel{d}{=} c_n X + d_n. \tag{4.28}$$

A key attribute that is evident immediately from the definition of an α-stable random variable is the property of closure under convolution of this family of random variables. This is apparent from the definition because it states that the linear combinations of two independent members of this family produces a transformed (convolved) random variable in the same family of distributions, with the same tail index α.

We will utilize this key property of α-stable random variables in the context of LDA models constructed with α-stable severity distributions to obtain closed-form analytic representations of the annual loss distribution. In general, there are a few special cases of parametric representations of α-stable models; these include two symmetric cases, the Gaussian distribution in which the tail index $\alpha = 2$ and the Cauchy case with tail index of 1 as well as the perfectly positively

skewed case of the Lévy distribution that we discuss in Example 4.1. In addition, there are two known cases of members of the family of α-stable models corresponding to $\alpha = 3/2$ and $\alpha = 2/3$ that can be represented by special functions given by the Holtsmark density (see Garoni and Frankel, 2002) and Whittaker functions (see Abramowitz and Stegun, 1965), respectively.

Remark 4.3 *The focus of modeling for the application of OpRisk severity models of a particular heavy-tailed loss process necessarily requires that the observed losses come from a severity distribution which only takes positive support. This can be achieved in three basic methods:*

1. *by transformation of the random variable;*
2. *by truncation of the distribution from which the random variate was drawn; or*
3. *by restriction of the parameter space to models with perfect positive skew that also produce positive support.*

We explore in detail the different approaches to restricting the α-stable family to positive support and then their incorporation into LDA OpRisk models in Chapter 6. We note here that the approach of transformation of a random variable such that its resulting severity distribution takes positive support will typically only be effective in practical settings in situations in which the distribution of the transformed random variable can be obtained as a function of the untransformed loss random variables severity distribution.

The approach of truncation is often very effective in practice; however, for many practical OpRisk modeling settings, it will require one to be able to quantify the truncation probability in order to normalize the resulting distribution on the truncated domain of positive support.

Remark 4.4 (Note on the Terminology of Truncation) *There are two different connotations or meanings applied to the use of the word truncation:*

1. *The first meaning involves the literal shortening or hard thresholding of the actual support of the distribution function and density functions. This will ensure strict positivity of the losses generated from such a truncated model.*
2. *The second meaning (often used in papers and books on this topic) is that of truncation of the spectral (Lévy) measure (soft truncation). This can result in distributions which are not necessarily taking only positive support for the loss random variables but they may be designed to have negligible (not strictly zero) probability of a negative loss outcome.*

The reason one may not choose to work in practice directly with the truncated case 1 above for the severity distributions in this classes of severity models is that for the α-stable family, this will typically result in a loss of the preservation of a particularly useful property of such models, that is, the closure under convolution of the severity distributions.

This closure under convolution means that given independent linear combinations of losses from such severity models, we can obtain the severity distribution of the linear combination in terms of the severity distribution of each loss. We will see that this is a particularly useful property to consider in OpRisk modeling in the compound process setting.

However, we will present a particular special case in Example 4.1 in which the truncation can be performed and will produce practically useful models in the α-stable family. These models will not only take positive support but will also be expressible analytically. In addition, we will consider the truncation setting for a second class of models which are less restrictive but will involve a series representation of the distribution and density, see Section 4.5.

EXAMPLE 4.1 Lévy Sub-Family of Severity Models

If $X \sim S(0.5, 1, \gamma, \delta; 0)$, this model specifies the sub-family of α-stable models with positive real support $x \in [\delta, \infty)$. The density and distribution functions are analytic and given, respectively, for $\delta < x < \infty$, by

$$f_X(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{1}{(x-\delta)^{3/2}} \exp\left(-\frac{\gamma}{2\,(x-\delta)}\right),$$

$$F_X(x) = \text{erfc}\left(\sqrt{\frac{\gamma}{2\,(x-\delta)}}\right).$$

The median is given by $\widetilde{\mu} = (\text{erfc}^{-1}(0.5))^2\,\widetilde{\gamma}/2$ and the mode $M = \widetilde{\gamma}/3$, where

$$\text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

This result follows from Nolan (2015, chapter 1, p. 5).

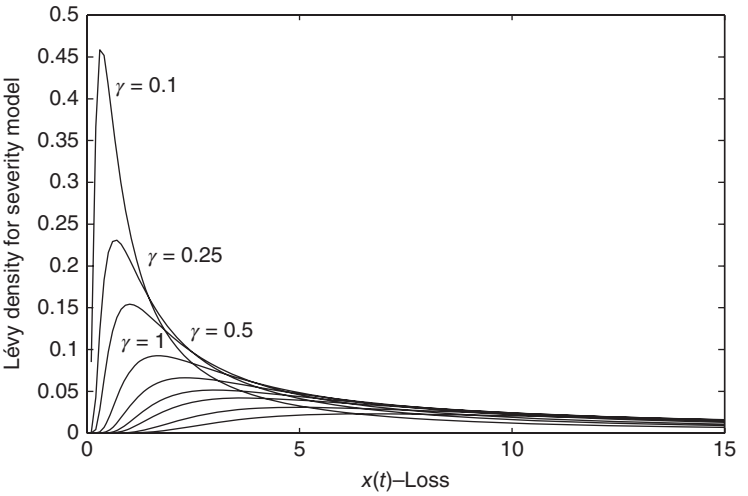

FIGURE 4.1 **Example of the Lévy severity model as a function of scale parameter, with location $\delta = 0$.**

Although we see that for the case of $\alpha = 1/2$ and $\beta = 1$, the resulting α-stable density is perfectly skewed; in general, for any α, setting $\beta = 1$ is not going to guarantee strictly positive support. That is, when discussing the notion of perfectly 'skewed' loss random variables from the α-stable severity models, this should also be carefully interpreted not necessarily as having a non-negative support (except for special parameter ranges). Instead as discussed in Marinelli and Rachev (2000), the total skewness left or right referred to by selection of the skewness parameter to one of the boundaries of its support is referring to the skew of the Lévy measure that characterizes its characteristic function representation (Figure 4.1). This will be discussed in detail later.

4.2.1 CHARACTERISATIONS OF α-STABLE SEVERITY MODELS VIA THE DOMAIN OF ATTRACTION

We will characterize below the family of all such distributions that satisfy the definitions for α-stable models in the equivalent Characterizations 1 and 2 above in terms of a third characterization known as the domain of attraction. We will denote such a family of models that satisfy this notion of domain of attraction of a stable law by the notation $\mho$. One generally does not consider situations in which random variable X is a point mass or degenerate as interesting in this context of defining $\mho$. That is, we will not discuss here classes of distributions for which linear combinations of loss random variables from distributions in this class, after scaling, converge to a Dirac mass (i.e. degenerate distribution).

We note that a generalization of Definition 4.7 for the α-stable family can be presented for any number of linearly combined random variables, to result in a generalized central limit theorem (CLT) result, as presented in Theorem 4.8. This result is fundamental to many areas of probability and statistics, and we note that Zolotarev (1986) summarized its importance well in the introduction of his influential text on univariate α-Stable models as follows:

> *"... It can be said without exaggerating that contemporary probability theory has acquired its independence, its authority as a mathematical discipline important for applications, and a large part of its arsenal of methods in the course of the solution of the problem of approximating distributions of sums of independent random variables..."*
>
> [V. M. Zolotarev]

Here we present an overview of the results regarding the generalization of the standard Gaussian CLT, which is assumed standard to most graduate students, extended to the wider setting of the α-stable CLT. There are several approaches one could take to introduce such a concept; we first introduce the approach which considers the domain of attraction of the limiting distribution for sequences of linear combinations of random variables. Then in characterizing the representations of the α-stable models that satisfy this domain of attraction, we consider the approach adopted in Zolotarev (1986) regarding the characteristic function representation.

First, we define the analogous concept to that presented in the extreme value theory (EVT) section relating to the maximal domain of attraction, in Definition 2.16. Here, instead we consider sequences of linear combinations of random variables, such as was discussed when talking about CLTs. The *domain of attraction* of a random variable comprised a linear combination of random variables is given in Definition 4.9. In particular, we will focus on the attraction domain of α-stable distributions in $\mho$, which was first studied in Gnedenko and Kolmogorov (1968), Ibragimov *et al.* (1971) and Petrov (1975).

Definition 4.9 (Domain of Attraction) *Consider a sequence of independent identically distributed random variables $X_i \sim F_X(x)$, $\forall i \in \{1, 2, \ldots, n\}$. Furthermore, consider all cases of distributions $F_X(x)$ for which there exists a suitable choice of constants A_n and B_n for $n = 1, 2, \ldots$, such that the distribution of the random variables X_i scaled and translated as follows*

$$\frac{\sum_{k=1}^{n} X_k - A_n}{B_n}, \quad n = 1, 2, \ldots, \tag{4.29}$$

converges weakly (in distribution), as $n \to \infty$, to a non-degenerate distribution function $G(x)$. Then denote the set of all such distributions $F_X(x)$ by $\mathcal{D}(G(x))$ which is the generic notation for the 'domain of attraction' for all such distributions for which a sequence of constance $A_n \in \mathbb{R}$ and B_n can be found to satisfy this weak convergence to distribution G.

In this chapter, we are particularly interested in the heavy-tailed cases and typically would be considering the space of sub-exponential distribution functions $F_X(x) \in \mathscr{F}$. Even in these cases where the integer moments of the random variables X_i in the sequence may not be finite, one can still achieve the aforementioned convergence in distribution to non-degenerate limiting distributions. It can be proven that not only does this convergence in distribution occur but as shown in Theorem 4.8, for suitable choices of constants a_n and b_n, this convergence will necessarily be to a stable law. In this case, we denote the domain of attraction for any stable family member by the notation $\mho$.

Theorem 4.8 (Characterization 3 of an α-Stable Severity Model) *Random variable X is stably distributed if it has a domain of attraction (denoted by $\mho$), such that there is a sequence of i.i.d. random variables $\{X_i; \forall i \in 1, 2, \ldots, n\}$ and sequences of positive random numbers $\{B_i \in \mathbb{R}^+; \forall i \in 1, 2, \ldots, n\}$ and $\{A_i \in \mathbb{R}; \forall i \in 1, 2, \ldots, n\}$ such that $\forall n \in \mathbb{J}$, the following convergence in distribution holds,*

$$\frac{\sum_{i=1}^{n} X_i}{B_n} + A_n \xrightarrow{d} X. \tag{4.30}$$

Notice that this result is dependent on the distributional properties of the sequence of random variables X_i and in particular does not preclude heavy-tailed random variables, so long as one can find the appropriate sequence of translation and scaling constants. This, therefore, generalizes the standard classical framework of finite moment random variables for X_i in which one would obtain some version of the classical CLT to the general case known as the stable limit theorem. We present an example of a sub-case of the stable limit theorem in Example 4.2.

EXAMPLE 4.2 Classical Central Limit Theorem Sub-case

In the most trivial of cases, consider i.i.d. $X_i \sim Normal(\mu, \sigma)$, then the result in Theorem 4.8 is a restatement of the closure under convolution of Gaussian random variables. The next well-known sub-case of the above result in Theorem 4.8 would correspond to one of the following two well-known results: Lindeberg Lévy CLT or the Lindeberg–Feller–Lyapunov CLT. Under these well-known results, there are additional restrictions imposed on the sequence of random variables, which further restrict the domain of attraction to that corresponding to the stable family member given by a Gaussian distribution for $G(x)$.

Lindeberg Lévy Central Limit Theorem. Suppose that $\{X_1, X_2, \ldots\}$ is a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathbb{V}\text{ar}[X_i] = \sigma^2 < \infty$. Then as $n \to \infty$, the random variables $B_n^{-1} \sum_{i=1}^{n} X_i + A_n$ for sequence choices $A_n = \sqrt{n}\mu$ and $B_n^{-1} = \sqrt{n}/n$ will result in the distributional convergence to a Gaussian distribution $G(x)$ with zero mean and variance σ^2, the family of all such distributions for random variables X_i are then in the domain of attraction of the Gaussian. These are clearly characterized for this sequence of A_n and B_n by finite second-moment distributions.

One can generalize this result by relaxing the conditions on the sequence of random variables $\{X_1, X_2, \ldots\}$ whilst still obtaining a very special case corresponding to $G(x)$

being a Gaussian distribution for the convergence. The following are the two popular relaxations:

- *Lyapunov Condition*. Suppose that the sequence $\{X_1, X_2, \dots\}$ is only comprised independent random variables (not necessarily identically distributed) with each random variable having finite mean μ_i and variance σ_i^2. Then defining for each $n \in \mathbb{J}^+$

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2, \tag{4.31}$$

  one can state the condition for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - \mu_i|^{2+\delta}\right] = 0. \tag{4.32}$$

  If this condition is satisfied, then the sequence of random variables in Theorem 4.8 will converge to a distribution $G(x)$ corresponding to the standard Gaussian distribution with $a_n = \sum_i^n \mu_i / s_n$ and $b_n = s_n$.
- *Lindeberg Condition*. Suppose that the sequence $\{X_1, X_2, \dots\}$ is only comprised independent random variables (not necessarily identically distributed) with each random variable having finite mean μ_i and variance σ_i^2. Then defining for each $n \in \mathbb{J}^+$

$$s_n^2 = \sum_{i=1}^{n} \sigma_i^2, \tag{4.33}$$

  one can state the condition for every $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_i)^2 \mathbb{I}_{|X_i - \mu_i| > \varepsilon s_n}\right] = 0. \tag{4.34}$$

  If this condition is satisfied, then the sequence of random variables in Theorem 4.8 will converge to a distribution $G(x)$ corresponding to the Gaussian distribution with $A_n = \sum_i^n \mu_i / s_n$ and $B_n = s_n$.

Of course, there are many other valid sequences of A_n and B_n constants that will satisfy convergence in Theorem 4.8 even if the moments of each random variable X_i in the sequence are not finite; this is the notion of the stable limit theorem; however, clearly, convergence to the distribution $G(x)$ will no longer be Gaussian, it will be an α-stable model. ■

This definition and Characterization 3 of the α-stable severity distributions via the domain of attraction of the stable family of models is complementary to the following definition of a stable random variable provided in Theorem 4.8. It can be further noted that this definition, whilst important and fundamental, is particularly useful only if one can describe the properties of the domains of attraction of stable laws. Progress on this was made for the case of the Normal distribution by Khintchine (1937), Feller (1945), Lévy and Borel (1954) and Lévy (1924) where they showed the result in Theorem 4.9.

Theorem 4.9 *A distribution function $F_X(x)$ belongs to the domain of attraction of a Normal law if and only if, as $x \to \infty$, the following convergence is true*

$$x^2 \frac{\int_{|y|>x} dF(y)}{\int_{|y|<x} y^2 dF(y)} \to 0.$$

This description of members of the Normal domain of attraction was further extended to stable laws with characteristic exponents $\alpha \in (0, 2)$ by Gnedenko and Kolmogorov (1968) and Doeblin (1940) as presented in Theorem 4.10.

Theorem 4.10 *A distribution function $F_X(x)$ belongs to the domain of attraction, $(\mathscr{D}(G(x)) \subseteq \mho)$, of a non-degenerate stable law $G(x)$ with characteristic exponent in the interval $\alpha \in (0, 2)$ if and only if, as $x \to \infty$, the following convergence is true*

$$\frac{F_X(-x)}{(1 - F_X(x) + F_X(-x))} \to \frac{c_1}{c_1 + c_2},$$

for some $c_1, c_2 \geq 0$ and $c_1 + c_2 > 0$, determined by $G(x)$. In addition, it is required that as $x \to \infty$

$$\frac{(1 - F_X(x) + F_X(-x))}{(1 - F_X(tx) + F_X(-tx))} \to t^\alpha,$$

for each constant $t > 0$.

Remark 4.5 *If one restricts the behaviour of the normalizing constants B_n in Definition 4.9, then one may recover narrower classes of distribution functions that satisfy the distributional convergence results specified in Definition 4.9. Furthermore, the Normal domain of attraction specified in Theorem 4.9 is formed by the set of distribution functions $F_X(x)$ that ensure convergence in distribution to stable distribution function $G(x)$, with exponent α, for a suitable choice of constants A_n, $c > 0$ and $B_n = cn^{-1/2}$, $n \in \{1, 2, \ldots, \}$. Finally, we note that the Normal domain of attraction coincides with the set of such non-degenerate distributions with finite variance.*

To conclude the specification of these results relating to the characterization of the α-stable severity models via their domain of attraction, one can consider the following sufficient conditions for the existence of the sequence of coefficients in Theorem 4.8, see Gnedenko and Kolmogorov (1968), Mijnheer and Mijnheer (1975), Aaronson and Denker (1998) and Marinelli and Rachev (2000) for details. Note that the results presented in Theorem 4.11 are identical to those discussed in Theorem 4.10.

Theorem 4.11 (Conditions for Membership of Stable Domain of Attraction) *Consider a sequence of i.i.d. loss random variables $\{X_i; \forall i \in 1, 2, \ldots, n\}$ with $X_i \sim F(x)$. For all $n \in \mathbb{J}$, one can find sequences of coefficients $\{A_i \in \mathbb{R}; \forall i \in 1, 2, \ldots, n\}$ and $\{B_i \in \mathbb{R}^+; \forall i \in 1, 2, \ldots, n\}$ that will satisfy the convergence in distribution given by*

$$\frac{\sum_{i=1}^n X_i}{B_n} + A_n \xrightarrow{d} X \tag{4.35}$$

for a random variable X which is stably distributed $X \sim S(\alpha, \beta, 1, 0; S1)$ (Nolans—$S1$ parameterizations) if the following necessary and sufficient conditions are satisfied for $F(x)$. The following function given by,

$$L(x) = x^\alpha (1 - F(x) + F(-x)) \tag{4.36}$$

must be slowly varying at infinity, and in addition, it should satisfy that

$$\lim_{x\to\infty} \frac{F(-x)}{1-F(x)+F(-x)} = \frac{1-\beta}{2}. \tag{4.37}$$

The sequence of coefficients a_n that will satisfy these conditions are given by

$$\lim_{n\to\infty} \frac{nL(a_n)}{A_n^\alpha} = \begin{cases} \Gamma(1-\alpha)\cos(\pi\alpha/2), & \alpha \in (0,1), \\ 2/\pi, & \alpha = 1, \\ \frac{\Gamma(2-\alpha)}{\alpha-|\cos(\pi\alpha/2)|}, & \alpha(1,2), \end{cases} \tag{4.38}$$

giving the sequence $A_n = n^{1/\alpha}L_0(n)$ with $L_0(n)$ a slowly varying function as $n \to \infty$. The sequence of coefficients B_n that will satisfy these conditions are given by

$$B_n = \begin{cases} 0, & \alpha \in (0,1), \\ nA_n \int \sin(x/A_n)\, dF(x), & \alpha = 1, \\ n\int x dF(x), & \alpha(1,2). \end{cases} \tag{4.39}$$

We will show that the property of infinite divisibility of α-stable distributions is instrumental in the development of results relating to analytic representations of LDA models under this family of heavy-tailed loss models. In addition, as demonstrated in the following sections, it allows one to specify the fourth characterization of the α-stable severity model according to representation of its characteristic function.

Definition 4.10 (Characterization 4 of an α-Stable Severity Model) *Random variable $X \sim S(\alpha, \beta, \gamma, \delta; 1)$ has a stable severity distribution if and only if it has a characteristic function that can be represented as follows:*

$$\mathbb{E}[\exp(\mathrm{i}\theta X)] = \begin{cases} \exp\{-\gamma^\alpha|\theta|^\alpha\left[1-\mathrm{i}\beta(\mathrm{sign}(\theta))\tan(\frac{\pi\alpha}{2})\right]+\mathrm{i}\delta\theta\} & \text{if } \alpha \neq 1 \\ \exp\{-\gamma|\theta|\left[1+\mathrm{i}\beta(\frac{2}{\pi})(\mathrm{sign}(\theta))\ln|\theta|\right]+\mathrm{i}\delta\theta\} & \text{if } \alpha = 1. \end{cases} \tag{4.40}$$

In the following sections, we will explain how this fourth characterization is obtained from several different perspectives, showing the links as we go between each of the four equivalent characterizations that have been provided.

4.3 Deriving the Properties and Characterizations of the α-Stable Severity Models

Having made these general stable distribution definitions, we can present the parameterized form of the univariate stable random variable. We adopt the notation of Nolan (2015), in which a univariate α-stable distribution is typically specified by four parameters: $\alpha \in (0, 2]$ determining the rate of tail decay (index of stability or characteristic exponent); $\beta \in [-1, 1]$ determining the degree and sign of asymmetry (skewness); $\gamma > 0$ the scale (under some parameterizations); and $\delta \in \mathbb{R}$ the location, see Lévy (1924). As discussed in Corollary 4.1, the parameter α is termed the characteristic exponent with small and large α implying heavy and light tails, respectively. We may now clarify the parametric examples in the stable family previously mentioned,

with respect to the tail index and the symmetry parameters; under a particular parameterizations to be presented later, one obtains the Gaussian ($\alpha = 2, \beta = 0$), Cauchy ($\alpha = 1, \beta = 0$) and Lévy ($\alpha = 0.5, \beta = 1$) distributions.

Except these special cases, in general, α-stable models admit no closed-form expression for the density which can be analytically evaluated pointwise. Therefore, statistical inference typically proceeds via the characteristic function, see discussions in Nolan (1998), Peters *et al.* (2010a) and Press (1972a). However, intractable to evaluate pointwise, importantly for OpRisk applications, simulation of random variates is very efficient (see Chambers *et al.* 1976).

We can now provide a mathematical representation of the family of univariate stable random variables, characterized by the characteristic function in Definition 4.18. Furthermore, we note that the scale parameter γ does not correspond to the standard deviation. We emphasize the parameterizations under consideration through the notation in which a random variable X is denoted as stably distributed with parameters $\alpha, \beta, \gamma, \delta$ under a particular parameterization k as specified by $X \sim S(\alpha, \beta, \gamma, \delta; k)$. When it is required to be clarified to avoid ambiguity, we may also utilize the notations for the stable model parameters given by α_k, β_k, γ_k and δ_k.

The following points summarize some of the key aspects we present in the following section:

- Introduce the $\mathbb{L}$-class of distributions and their relationship to the properties of α-stable models such as unimodality and characteristic function representation under the Lévy–Khinchin form.
- Basic steps in going from the Lévy–Khinchin characteristic function representation of an α-stable model to the parameterizations utilized widely in practical settings to specify the characteristic function of an α-stable model. More precisely, how to go from the Lévy measure and spectral function representation to the closed-form analytic representation of Zolotarev's A-type parameterizations of the characteristic function.
- Specification of different α-stable characteristic function parameterizations and their relationships: Zolotarev's A, M, B, W, C and E types as well as Nolan's S0 and S1 types.

To proceed, we first define the Stieltjes integral which will allow us to present the stable characteristic function.

Definition 4.11 (Stieltjes Integral) *The Stieltjes integral of a real function $f(x)$ relative to a real function $G(x)$ of bounded variation on an interval $[a, b]$ is defined, analogously to the Riemann integral, as a limit of a sum of terms $f(a_i)[G(x_i) - G(x_{i-1})]$ taken as partitions of the interval shrink. Furthermore, if $G(x)$ is differentiable with $G'(x) = g(x)$, then the Stieltjes integral can be expressed in terms of the Riemann integral (if it exists) according to*

$$\int_a^b f(x)\,dG(x) = \int_a^b f(x)g(x)\,dx. \tag{4.41}$$

Having defined in a very simple manner, this form of integration, we may now, define the characteristic function of a univariate random variable as the expectation operator on a complex exponential transformation of the random variable as given in Definition 4.12.

Definition 4.12 (Characteristic Function) *The definition of the characteristic function or the Fourier transform of a univariate random variable X is given by*

$$\Phi_X(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta X)] = \int_{-\infty}^{\infty} \exp(\mathrm{i}\theta x) dF_X(x). \tag{4.42}$$

4.3.1 UNIMODALITY OF α-STABLE SEVERITY MODELS

In this section, we show that distributions in the domain of attraction of a stable limit $\mho$ that satisfy the convergence in Theorem 4.8 will all be univariate. We first start by noting the following general definition of unimodality of a distribution given in Definition 4.13.

Definition 4.13 (Unimodality of a Distribution) *A distribution function $F(x)$ is unimodal with a mode m if $F(x)$ is convex for $x < m$ and concave for $x > m$.*

A useful property that is sufficient for membership of a distribution in the class of unimodal models is based on the strict log-concavity of the density (when it exists). In fact, the property of log-concavity is crucial to the membership of the unimodal class of distributions and in addition tells the modeller useful additional information. For instance, this sufficient condition has important implications for sampling form such distributions, especially via adaptive rejection type mechanisms, which will have practical application if one adopts an α-stable model with truncation, as will be discussed later. However, in the case of the α-stable distribution, it is non-trivial to verify the log-concavity of the density, because there exists no closed-form expression for representation of the density in a simple manner that would allow, for all possible parameter values, the verification of the log-concavity of the density.

Instead, progress on this result was made in an alternative manner discussed briefly later. In developing an understanding of the representation of the α-stable family of severity distributions, as specified by the characteristic function representation, it is informative to first define the class of $\mathbb{L}$ distributions, specified in Definition 4.14, see Gnedenko and Kolmogorov (1968).

Definition 4.14 ($\mathbb{L}$ Class of Distributions) *A distribution function F is in the class of $\mathbb{L}$ distribution functions if there exists a sequence of independent random variables $\{X_n\}_{n\geq 1}$ such that for suitably chosen constants $B_n > 0$ and A_n, the distribution functions of the sums given by*

$$Y_n = B_n^{-1} \sum_{k=1}^{n} X_k - A_n, \tag{4.43}$$

converge to the distribution F and the random variables

$$X_{nk} = \frac{X_k}{B_n} \tag{4.44}$$

are asymptotically constant as $n \to \infty$.

Clearly, the definition of the class of L-distributions leads one to consider further the notion of self-decomposability of random variables as detailed in Definition 4.5. Self-decomposable distributions are given as limits of normalized sums of independent

random variables such as the geometric stable family studied in the insurance section in (Cruz *et al.* 2015, chapter 15). Recently, there has also been interesting developments in semi-self-decomposability, see discussions in Maejima and Naito (1998). This notion of self-decomposable distributions naturally leads one to the L-class of distributions as was clear from the definition of this class.

Remark 4.6 *Note that from this definition of the class of $\mathbb{L}$ distributions, a distribution F in this class will always be infinitely divisible. Furthermore, α-stable models are known to be members of the $\mathbb{L}$-type distribution class.*

The necessary and sufficient conditions for membership of a distribution in the class of $\mathbb{L}$ distributions was specified by Lévy (1939) and Lévy and Borel (1954) and given in Proposition 4.1.

Proposition 4.1 (Distributional Membership of $\mathbb{L}$ Class) *A distribution function $F(x)$ will belong to the class of $\mathbb{L}$ distributions if the following necessary and sufficient conditions are satisfied for its log-characteristic function $\ln(\Phi_X(\theta))$. The characteristic function of the random variable X with distribution $F(x)$ can be represented in the following manner*

$$\begin{aligned}\ln(\Phi_X(\theta)) = \mathrm{i}\gamma\theta - \frac{1}{2}\sigma^2\theta^2 &+ \int_{-\infty}^{0_-}\left(\exp(\mathrm{i}\theta u) - 1 - \frac{\mathrm{i}\theta u}{1+u^2}\right)|u|^{-1}l(u)\,du \\ &+ \int_{0_+}^{\infty}\left(\exp(\mathrm{i}\theta u) - 1 - \frac{\mathrm{i}\theta u}{1+u^2}\right)u^{-1}k(u)\,du, \end{aligned} \tag{4.45}$$

with functions $k(u) \geq 0$ and $l(u) \geq 0$, $-l(u)$ and $k(u)$ non-increasing and satisfying the two conditions

$$\begin{aligned}\int_{-1}^{0_-}|u|l(u)\,du + \int_{0^+}^{1}uk(u)\,du < \infty,\\ \int_{-\infty}^{-1}|u|^{-1}l(u)\,du + \int_{1}^{\infty}u^{-1}k(u)\,du < \infty.\end{aligned} \tag{4.46}$$

Remark 4.7 *In the following sections, when introducing the characterization of the α-stable family of models with regard to their characteristic function, we will go into greater detail regarding the representation of the log-characteristic function presented earlier, which is widely known as the Lévy–Khinchin formula or representation.*

The observation that α-stable models are a sub-family of the $\mathbb{L}$ class of distributions proved very useful in characterizing several important properties of α-stable models. The first of these properties is presented in Theorem 4.12, see Yamazato (1978, theorem 1) and Chernin and Ibragimov (1959).

Theorem 4.12 (Unimodality of the $\mathbb{L}$ Class of Distributions) *All infinitely divisible distribution functions $F(x)$ in the class of $\mathbb{L}$ distributions are unimodal.*

The verification of the following corollary regarding membership of the α-stable family of models in the $\mathbb{L}$ class of distributions is discussed in the following section.

Corollary 4.2 (Unimodality of the α-Stable Family of Models) *The class of α-stable distributions characterized by domain of attraction $\mho$ are also members (a sub-family) of the class of $\mathbb{L}$ distributions. Therefore, all α-stable models will be strictly unimodal.*

A good overview of properties of unimodality of such distributional classes is provided in the discussion on general unimodality in Olshen and Savage (1969). To understand further these results on unimodality, it should be noted that they came from a sequence of works we briefly outline later. It was proven in Wintner (1938) and Wintner (1956) that all *symmetric* stable laws are unimodal, a useful property to know with certainty about these severity models, no matter what parameter values are considered or estimated in the symmetric case, especially when developing this family of severity models in OpRisk. Know that such a feature will hold is beneficial to the understanding and interpretation of the process and its attributes once fitted to real data. However, in OpRisk, it is typically going to be the case that non-symmetric distributions will be of relevance, so it would be important to obtain knowledge of the potential unimodality of non-symmetric distributions in this family.

This result was then generalized beyond the symmetric case and was shown to also hold for asymmetric stable models. To achieve this generalization, it was shown in Yamazato (1978) that this unimodality result for symmetric stable laws could be significantly generalized for the class of $\mathbb{L}$ distributions and, therefore, for all α-stable models including asymmetric models. Yamazato (1978) was able to prove that all infinitely divisible distributions of class $\mathbb{L}$ were unimodal, implying that all α-stable distributions will be unimodal distributions which includes the sub-family of asymmetric α-Stable models.

Yamazato's (1978) more general approach was based on a proof which relied on the fact that any α-stable law can be written as a limit of distributions which are known to be unimodal and, therefore, itself will be unimodal. This result from Yamazato (1978) itself was based on relaxing an earlier result of Ibragimov (1956) which relied on the concept of strong unimodality of distribution functions as given in Definition 4.15.

Definition 4.15 (Strong Unimodality) *A strongly unimodal distribution is one in which under convolution with every unimodal distribution function, it remains unimodal.*

4.3.2 RELATIONSHIP BETWEEN $\mathbb{L}$ CLASS AND α-STABLE DISTRIBUTIONS

Next we discuss how knowledge of the fact that α-stable models are members of the class of $\mho \subset \mathbb{L}$ distributions allows us to represent such models via generic characteristic function representation due to the Lévy–Khinchin. To validate this membership, we will make explicit the functional forms for the 'Lévy triplet' in the α-stable family of distributions, showing that they can be represented in a manner that satisfies membership of the $\mathbb{L}$ class given knowledge of the Lévy measure for an α-stable model.

Then we will relax the presumed knowledge of the α-stable models, in particular no longer assuming knowledge of the Lévy spectral measure, instead relying only on the fundamental properties of closure under convolution and infinite divisibility to obtain the required membership for representation in the class of $\mathbb{L}$ distributions. This will involve a brief summary of the key steps involved in the detailed derivation of such results provided in Zolotarev (1986),

illustrating to the reader the key stages of obtaining such a fundamental result. For a detailed mathematical discussion on this result, the interested reader is referred to Zolotarev (1986) and Samorodnitsky and Taqqu (1997).

As noted above in Proposition 4.1, it is sufficient to show membership of a distribution in the $\mathbb{L}$ class of distributions if the log-characteristic function can be shown to have the representation that has become known as the Lévy–Khinchin form. At this point, we will first introduce a slightly different representation of the class of $\mathbb{L}$ distributions, via the Lévy measure characterization of the characteristic function, see Feller (1966, p. 542). This is simply a convenient restatement of the previous representation which is in terms of the spectral measure $W(du)$, the analogous quantity to the functions $l(u)$ and $k(u)$ presented previously.

Definition 4.16 (Lévy–Khinchin Representation) *Given a distribution function F which is infinitely divisible, then the logarithm of its characteristic function admits the Lévy–Khinchin representation given generally by the canonical representation*

$$\Phi_X(\theta) = \exp(\psi(\theta)), \quad \theta \in \mathbb{R}, \tag{4.47}$$

where the exponent of the characteristic function $\psi(\theta)$ is given by

$$\psi(\theta) = \mathrm{i}a_g\theta - \tfrac{1}{2}\sigma^2\theta^2 + \textstyle\int_{\mathbb{R}\setminus\{0\}}(e^{\mathrm{i}\theta u} - 1 - \mathrm{i}\theta g(u))\, W(du), \tag{4.48}$$

for some truncation function $g : \mathbb{R} \to \mathbb{R}$ which is measurable and satisfies for all θ the condition

$$\int |\exp(\mathrm{i}\theta u) - 1 - \mathrm{i}\theta g(u)| W(du) < \infty, \tag{4.49}$$

with $a_g \in \mathbb{R}$ depending on the choice of truncation function and $\sigma \geq 0$ and the Lévy measure W which satisfies

$$\int_{\mathbb{R}\setminus\{0\}} \min\{u^2, 1\} W(du) < \infty. \tag{4.50}$$

Remark 4.8 (Lévy Triplets) *Note that if observing losses continuously over time, the resulting sub-class of α-stable process can be represented according to the Lévy process. Furthermore, this is uniquely defined by the Lévy triplet (a_g, σ^2, W) corresponding to a drift, a diffusion component and a jump component. This will be particularly relevant where tempered stable and generalized tempered stable processes are introduced under such a representation.*

As noted, there are many choices one could make for the truncation function in the Lévy–Khinchin representation of the characteristic function and it is important to recognize that these choices will not affect the spectral measure $W(du)$ or the diffusion component σ^2, they only affect the drift component a_g.

Remark 4.9 *Another equivalent condition that the truncation function $g(u)$ should satisfy is that it can be any function of the form $g : \mathbb{R} \to \mathbb{R}\setminus\{0\}$ that satisfies the condition that*

$$\int_{u\neq 0} \left| \frac{g(u) - u}{u^2} \right| du < \infty. \tag{4.51}$$

It is then clear that depending on the choice made for the function $g(u)$ in the specification of the Lévy–Khinchin representation, one will obtain different representations only for the volatility component a_g in the Lévy triplet. Popular choices of the function $g(u)$ include the forms (not an exhaustive list):

- *Choice 1*: $\dfrac{u}{u^2+1}$
- *Choice 2*: $u\mathbb{I}_{|u|<1}$
- *Choice 3*: $\dfrac{u}{\max\{u,1\}}$
- *Choice 4*: $\sin(u)$
- *Choice 5*: $u\mathbb{I}_{u\in[0,1]} + u(2-u)\mathbb{I}_{u\in(1,2]}$.

One can move between different functions by simply modifying the drift component a_g, where it is convention to specify these choices relative to the reference denoted by a which corresponds to Choice 2. One can then move between the other choices relative to Choice 2 as follows:

$$a_g = a + \int_{-\infty}^{\infty} \left[g(u) - u\mathbb{I}_{|u|<1} \right] W(du). \tag{4.52}$$

To relate the role of self-decomposability of a loss random variables distribution to the Lévy–Khinchin representation of its characteristic function, we note that all self-decomposable distributions (laws) are characterized by their Lévy density of the form given in Equation 4.53, for some function $h(u)$ which is increasing for $u<0$ and decreasing for $u>0$. This function $h(u)$ is known as its self-decomposability characteristic.

$$W(du) = \frac{h(u)}{|u|}\mathbb{I}_{u<0} + \frac{h(u)}{u}\mathbb{I}_{u>0}. \tag{4.53}$$

Following this general Lévy–Khinchin characteristic function representation of infinitely divisible distributions in class $\mathbb{L}$, we consider restricting our attention to the sub-class of α-stable members. One can then show that given this general characterization of the Lévy–Khinchin representation of members of class $\mathbb{L}$ log-characteristic functions, the class of α-stable distributions have characteristic functions that satisfy this representation. One way to proceed in verifying this result is to assume that one has the knowledge of the appropriate form of the Lévy spectral measure characterizing the α-stable family. For instance, the discussion in Janicki and Weron (1994a) and references therein provide the α-stable Lévy measure functional form given in Definition 4.17.

Definition 4.17 (α-Stable Lévy Measure) *The α-stable Lévy measure is given in Janicki and Weron (1994a, section 2.3, 1994b)*

$$W(du) = \frac{P}{u^{1+\alpha}}\mathbb{I}_{(0,\infty)}(u)\,du + \frac{Q}{|u|^{1+\alpha}}\mathbb{I}_{(-\infty,0)}(u)\,du, \tag{4.54}$$

with non-negative numbers P and Q. It will also be useful to consider an equivalent alternative definition provided in Zolotarev (1986, theorem C.1, Equation I.10) which defines the spectral function by

$$W(u) = \begin{cases} -C_1 u^{-\alpha}, & \text{if } u>0, \\ C_2(-u)^{-\alpha}, & \text{if } u<0, \end{cases} \tag{4.55}$$

for C_1, C_2 non-negative numbers and $\alpha \in (0, 2)$.

Then if one chooses the functional form for the characteristic function of an α-stable model given generically by

$$\mathbb{E}[\exp(\mathrm{i}\theta X)] = \begin{cases} \exp\left(\mathrm{i}a\theta - \sigma^2\theta^2\right), & \text{if } \alpha = 2, \\ \exp\left(\mathrm{i}a\theta - \int_{\mathbb{R}\setminus\{0\}} \Psi(\theta, u) W(du)\right), & \text{if } 0 < \alpha < 2, \end{cases} \tag{4.56}$$

for $a \in \mathbb{R}$ and $\sigma > 0$ and after substitution of the α-stable Lévy measure in Definition 4.17 and selection of the function $\Psi(\theta, u)$ according to the form

$$\Psi(\theta, u) = \exp(\mathrm{i}\theta u) - 1 - \frac{\mathrm{i}\theta u}{1 + u^2}, \tag{4.57}$$

one obtains the required representation of the α-stable model in a form that satisfies membership of the class of $\mathbb{L}$ distributions.

In addition, this representation then allows one to rearrange the characteristic function, making appropriate change of variables to obtain standard representations of the α-stable characteristic function commonly used in practice, see, for instance, discussions in Nolan (1997a). Hence, if we assume knowledge of the Lévy spectral measure of the α-stable family, we may conclude that α-stable distributions are a sub-class of $\mathbb{L}$ distributions; this results in the following fourth-equivalent characterization. If we do not want to assume this information and instead wish to utilize the fundamental definitions of an α-Stable distribution given in Theorem 4.8, one can still derive the result for membership of α-stable distributions in the class of $\mathbb{L}$ distributions. This is more involved, so a brief sketch of the required steps is provided in the following section.

4.3.3 FUNDAMENTALS OF OBTAINING THE α-STABLE CHARACTERISTIC FUNCTION

A more detailed instructive approach to obtaining this representation of the characteristic function of an α-stable distribution was provided by Zolotarev (1986) or indeed in other monographs such as Gnedenko and Kolmogorov (1968). The approach of Zolotarev starts from the perspective of defining sequences of centred random variates and looks at the properties of their limiting distributions asymptotically under certain probabilistic conditions on the maximum values they can achieve. In the following brief section, we recapitulate this approach.

Under the approach of Zolotarev (1986), one first considers a sequence of loss random variables, denoted by $\{X_{nj} \sim F_{nj}, j = 1, 2, \ldots, k_n\}$ for $n = 1, 2, \ldots$, utilized to construct the centered sums given by

$$Z_n = X_{n1} + X_{n2} + \cdots + X_{nk_n} - A_n \sim F_n, \tag{4.58}$$

for known translation A_n. Next it is assumed that each component of the linear combination is uniformly infinitesimal. This assumption is characterized by the following limit on the maximum probability of each term exceeding an ε threshold and the probability that this exceedance goes asymptotically to zero as the number of terms in the sequence increases, according to

$$\lim_{n\to\infty} \max\left\{\mathbb{P}\mathrm{r}\left[|X_{nj}| > \varepsilon\right] : j = 1, \ldots, k_n\right\} = 0. \tag{4.59}$$

Remark 4.10 *A special case of the sequence of random variables considered earlier (typically considered in many practical settings) would correspond to the case where for each $n \in \mathbb{J}^+$, one defines*

$$Z_n = \frac{1}{B_n}(X_1 + X_2 + \cdots + X_n) - A_n \sim F_n, \tag{4.60}$$

where $B_n \to \infty$ as $n \to \infty$. As noted in Zolotarev (1986), the property in Equation 4.59 will hold for the above sequence whenever F_n converges in distribution to a non-degenerate distribution G.

Now define the set $\mathcal{G}$ to be the set of all distribution functions G that correspond to the weak limits of distribution functions F_n as $n \to \infty$, then one can show that $\mathcal{U} \subset \mathcal{G}$. Note that we have already denoted the domain of attraction of the α-stable distributions by $\mathcal{U}$ that we state now is a subset of the domain of attraction of these distributions in the sequence F_n, denoted by $\mathcal{G}$.

Under such a framework, we can use the following fundamental representation of such distribution functions G to be members of $\mathcal{G}$ which when we consider the stable members will correspond to $\mathcal{D}(G)$ being given by $\mathcal{U}$, given by Theorem 4.13 from Zolotarev (1986, p. 4, theorem A). This result is known as the canonical form of a representation of the α-stable distributions characteristic function in terms of the spectral function W.

Theorem 4.13 *A distribution function G belongs to the set $\mathcal{G}$ if and only if the log-characteristic function can be written in the Lévy–Khinchin representation form in Definition 4.16.*

As a result of this theorem, we may now focus purely on the Lévy–Khinchin representation of the characteristic function and identify conditions for the Lévy triplet required to produce the domain of attraction $\mathcal{G}$ of which we are particularly interested in the sub-class of the α-stable domain of attraction for such sequences of random variables. It was then shown in Gnedenko (1983), Petrov (1975), Gnedenko and Kolmogorov (1968) and Gnedenko (1939) that the required conditions on the Lévy triplet to produce the $\mathcal{G}$ domain of attraction are given in Theorem 4.13. The results in this theorem rely upon two moment conditions for the representation of normalizing constant sequence mean and variance when restricted to a finite interval related to the Lévy spectral measure, given as Condition 1 and Condition 2. These two conditions when combined with the representation from Theorem 4.13 are sufficient to obtain the weak convergence of the distributions F_n to $G \in \mathcal{G}$.

Remark 4.11 *We note that in the special case that one consider the simple sequence defined in Equation 4.60, one obtains the weak convergence of the distributions F_n to G where G is one of the family members which forms the limiting distribution for the domain of attraction $\mathcal{U} \subset \mathcal{G}$, the α-stable family.*

Condition 1. Define the sequence $a_n(y)$, for a threshold $y \in \mathbb{R}^+$ in which y and $-y$ are continuity points of the spectral function $W(x)$, according to the expected value of the linear combination of an additive functional of random variates in the sequence $\{X_{nj} : j = 1, 2, \ldots, k_n\}$. Take the additive functional as the indicator function which restricts the expectation on each member of the sequence to a set corresponding to the event that the absolute value of the random variable in the sequence exceeds the threshold y according to

$$a_n(y) = \sum_{j=1}^{k_n} \mathbb{E}[X_{nj}\mathbb{I}(|X_{nj}| < y)]. \tag{4.61}$$

This expectation for threshold y, given by $a_n(y)$, will then be used in the construction of the centring coefficient A_n in the sequence $\{Z_n\}_{n\in\mathbb{J}^+}$.

Condition 2. Furthermore, define the variance of the additive function, on which the expectation was applied to the sequence of random variables $\{X_{nj} : j = 1, 2, \ldots, k_n\}$, according to

$$\sigma_n^{\varepsilon} = \sum_{j=1}^{k_n} \mathbb{V}\mathrm{ar}(X_{nj}\mathbb{I}(|X_{nj}| < \varepsilon)), \tag{4.62}$$

for some threshold $\varepsilon > 0$.

Note, in the case that the distribution G turns out to be a Gaussian random variable, these conditions are similar in spirit to the Lindeburg or Lyapanuv conditions discussed in Example 4.2.

Under these two conditions, Zolotarev (1986, theorem B, p. 4.) states the following weak convergence results for the distributions F_n to converge to $G \in \mathcal{G}$ asymptotically. This will then form the basis of the weak convergence results for α-stable models when we consider the subset $\mathcal{U} \subset \mathcal{G}$.

Theorem 4.14 *Consider G as a distribution function in the class of $\mathcal{G}$ with a canonical characteristic function representation given by Theorem 4.13, uniquely specified by the triple a, b and spectral function W. Next consider the sequences of random variables $\{X_{nj} : j = 1, 2, \ldots, k_n\}$ and the resulting linear combinations constructed from these sequences given by*

$$Z_n = X_{n1} + X_{n2} + \cdots + X_{nk_n} - A_n \sim F_n, \tag{4.63}$$

and select the centering constants A_n according to

$$A_n = a_n(y) - a - \int_{|u|<y} u \; dW(u) + \int_{|u|\geq y} u^{-1} \; dW(u). \tag{4.64}$$

Then the distribution functions F_n converge weakly to G as $n \to \infty$ if and only if the following two conditions are met:

1. $\lim_{n\to\infty} \sum_{j=1}^{k_n} \left(F_{nj}(x) - \frac{1}{2} - \frac{1}{2}\mathrm{sgn}(x)\right) = W(x)$; *and*
2. $\lim_{\varepsilon\to 0}\lim_{n\to\infty} \sup \sigma_n^{\varepsilon} = \lim_{\varepsilon\to 0}\lim_{n\to\infty} \inf \sigma_n^{\varepsilon} = 2b$.

This, therefore, provides an alternative, perhaps more pedagogical, route to achieving the representation of the characteristic function of the α-stable laws, which is eventually the same representation achieved in Equation 4.56. To see this, consider the simplest version of the sequence defined in Equation 4.58 is the sequence of linearly normalized sums of i.i.d. random variables given in Equation 4.29 above, where a_n and $b_n > 0$ are real constants.

To insure that the condition in Equation 4.59 is satisfied in the sequence defined in Equation 4.58, it is sufficient to have the sequence $b_n \to \infty$ as $n \to \infty$. This will be true whenever the distribution functions F_n converge weakly to a non-degenerate distribution G. Furthermore, this provides the definition of a stable distribution $G \in \mathcal{U}$. That is, a distribution G is α-stable if it occurs as the weak limit of the distribution functions F_n as $n \to \infty$ for sequences given in Equation 4.58.

The α-stable laws simply characterize the set of all such distributions that satisfy this weak limit, and we will denote them according to either a subset of the domain of attraction $\mathscr{D}(G)$ or by $\mho$. As noted in Zolotarev (1986), there are numerous criteria that can be considered to characterize the members of the α-stable family, some of which have been provided in the previous subsection.

Remark 4.12 *To this point, we have characterized the α-stable family of severity distributions according to the domain of attraction corresponding to the weak convergence of sequences of appropriately scaled and normalized sums of loss random variables. This was done in the most general manner through the justification of the Lévy–Khinchin representation under two alternative approaches:*

- *The first appealed to properties of the characterization of the L-Class of distributions and presumed its knowledge of the α-Stable families membership in this class to obtain the Lévy–Khinchin characterization.*
- *The second approach (directly above) was more general and motivated the Lévy–Khinchin representation from first principles culminating in Theorem 4.13 for the canonical form of representation of the α-stable laws, for which two conditions were provided that were sufficient for the sequences of random variables and normalizing constants to satisfy for this representation to hold.*

Working with the Lévy–Khinchin representation of the α-stable laws is typically not practically useful for applications. It is, therefore, more common to work with one of the several different parameterizations. Next we first state a reparameterized form of the α-stable distribution family—the A-Type parameterizations of Zolotarev, and we demonstrate the basic steps to go from the Lévy–Khinchin representation to this parameterizations. Then we show the relationship between this parameterization and several other parameterizations used in practical settings.

Two additional properties of interest, corresponding to results $I.8$ and $I.10$ in Zolotarev (1986), are presented, followed by Theorem C.2 of Zolotarev (1986) which presents the A-type parameterizations of an α-stable characteristic function based on these two properties. In detailing this parameterizations, we will sketch out the process of taking the Lévy–Khinchin representation of the α-stable characteristic function in Theorem 4.13 (analogous to that provided in Equations 4.56 with spectral Lévy measure given in Definition 4.17) and achieving the representation Zolotarev's A-type representation via contour integration. This can then be linked to all other parametric parameterizations in use.

In Lemma 4.2, we characterize a useful criterion that members of $G \in \mathscr{G}$ must satisfy to correspond to the stable distributions characterized by the domain $\mathscr{D}(G) = \mho$.

Lemma 4.2 *A distribution function G belongs to the α-stable family, characterizing a sub-family of the domain of attraction $\mho \subset \mathscr{G}$ if and only if it has the property that for any positive numbers b_1 and b_2, there exists a positive number b and a real number a such that the following distributional convolution holds*

$$G\left(\frac{x}{b_1}\right) \star G\left(\frac{x}{b_2}\right) = G\left(\frac{x-a}{b}\right). \tag{4.65}$$

From this Lemma 4.2, we can now state that a condition of membership for any distribution G to make up a subset of $\mho$ is that it takes the representation of its characteristic function given in Theorem 4.15 for the A-type parameterizations of Zolotarev (1986).

Theorem 4.15 (Zolotarev's A-Type α-Stable Characteristic Function) *A non-degenerate distribution satisfies G characterizes a sub-family of the limiting distributions for $\mho$ if and only if it has log-characteristic function given by*

$$\ln\left(\Phi_X(\theta)\right) = \lambda(\mathrm{i}\theta\gamma - |\theta|^{\alpha} + \mathrm{i}\theta\omega_A(\theta,\alpha,\beta)), \quad \theta \in \mathbb{R}, \tag{4.66}$$

where the real parameters vary within the limits $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$ and $-\infty < \gamma < \infty$, $\lambda > 0$ and

$$\omega_A(\theta,\alpha,\beta) = \begin{cases} |\theta|^{\alpha-1}\beta\tan\left(\frac{\pi\alpha}{2}\right) & \text{if } \alpha \neq 1, \\ -\beta\left(\frac{2}{\pi}\right)\ln|\theta| & \text{if } \alpha = 1. \end{cases} \tag{4.67}$$

Remark 4.13 *This parameterization is typically not utilized in practice because of the feature it possesses which involves a widely discussed discontinuity in the characteristic function and the distribution functions G. That is, the A-type Zolotarev parameterizations is not continuous with respect to the parameters, because, clearly, the function $\omega_A(\theta,\alpha,\beta)$ has discontinuities at all points of the form $\alpha = 1, \beta \neq 0$.*

Fortunately, this representation can be reparameterized to overcome this difficulty as discussed in the following section.

4.3.4 FROM LÉVY–KHINCHIN'S CANONICAL REPRESENTATION TO THE α-STABLE CHARACTERISTIC FUNCTION PARAMETERIZATIONS

In this section, we will demonstrate how to derive the Zolotarev's A-Type parameterizations, starting from the canonical Lévy–Khinchin representation provided in Theorem 4.13 and sketching the steps to arrive at the result in Theorem 4.15, as performed in detail in Zolotarev (1986, p. 9–11). To start with, we will show the relationship between Zolotarev's A-Type parameterizations that relates to Characterization 4 and the alternative representation provided in Characterization 2.

4.3.4.1 From Characterization 4 to Characterization 2. Consider a loss random variable $X \sim S(\alpha,\beta,\lambda,\gamma;A)$ and n independent copies denoted by $X_1, X_2, \ldots, X_n$. If Characterization 4 is considered (under the Zolotarev A-Type parameterisations), one can, therefore, represent the characteristic function according to the following expression

$$\Phi_X(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta X)] = \exp[\lambda(\mathrm{i}\theta\gamma - |\theta|^{\alpha} + \mathrm{i}\theta\omega_A(\theta,\alpha,\beta))], \quad \theta \in \mathbb{R}. \tag{4.68}$$

To show this is equivalent to the Characterization 2, we first consider the joint characteristic function of the sum of the n losses given for the partial sum of independent copies

$$S_n = \sum_{i=1}^{n} X_i, \tag{4.69}$$

given by the form

$$\begin{aligned}\Phi_{S_n}(\theta) &= \mathbb{E}[\exp(\mathrm{i}\theta S_n)] \\ &= \mathbb{E}[\exp(\mathrm{i}\theta(X_1 + X_2 + \cdots + X_n))] \\ &= \exp[n\lambda(\mathrm{i}\theta\gamma - |\theta|^\alpha + \mathrm{i}\theta\omega_A(\theta,\alpha,\beta))] \\ &= \begin{cases} \exp\left[n\lambda\left(\mathrm{i}\theta\gamma - |\theta|^\alpha + \mathrm{i}|\theta|^{\alpha-1}\beta\tan(\pi\alpha/2)\right)\right], & \text{if } \alpha \neq 1, \\ \exp[n\lambda\left(\mathrm{i}\theta\gamma - |\theta|^\alpha - \mathrm{i}\theta\beta\left(2/\pi\right)\ln|\theta|\right)], & \text{if } \alpha = 1. \end{cases}\end{aligned} \tag{4.70}$$

Then one also has for the scaled and translated random variable $c_n X + d_n$ the representation

$$\begin{aligned}\Phi_{c_n X + d_n}(\theta) &= \exp(\mathrm{i}\theta d_n)\mathbb{E}[\exp(\mathrm{i}\theta c_n X)] \\ &= \exp(\mathrm{i}\theta d_n)\exp[\lambda(\mathrm{i}c_n\theta\gamma - |c_n\theta|^\alpha + \mathrm{i}c_n\theta\omega_A(c_n\theta,\alpha,\beta))] \\ &= \begin{cases} \exp(\mathrm{i}\theta d_n)\exp\left[\lambda\left(\mathrm{i}c_n\theta\gamma - |c_n\theta|^\alpha + \mathrm{i}c_n\theta|c_n\theta|^{\alpha-1}\beta\tan(\pi\alpha/2)\right)\right], \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha \neq 1 \\ \exp(\mathrm{i}\theta d_n)\exp[\lambda\left(\mathrm{i}c_n\theta\gamma - |c_n\theta|^\alpha - \mathrm{i}c_n\theta\ln|c_n\theta|\beta\left(2/\pi\right)\right)], \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha = 1. \end{cases}\end{aligned} \tag{4.71}$$

Then the equivalence between the two characterizations becomes clear when one makes the selection for the coefficients

$$\begin{aligned} c_n &= n^{1/\alpha}, \\ d_n &= \lambda\gamma\left(n - n^{1/\alpha}\right). \end{aligned} \tag{4.72}$$

4.3.4.2 From Characterization 2 to Characterization 4. Next we highlight the key stages in making the equivalence in the opposite direction by starting with Characterization 2 and achieving Characterization 4. The following is a brief overview to demonstrate the six steps required (for full details, see Zolotarev 1986).

Step 1: Consider the sequence of random variables

$$Z_n = \frac{1}{B_n}(X_1 + X_2 + \cdots + X_n) - A_n \sim F_n, \tag{4.73}$$

that satisfy Theorem 4.13, and the Condition 1 and Condition 2 required for Theorem 4.13 to hold. Then one can characterize the spectral measure of the sequence generically by

$$\lim_{n\to\infty}\sum_{j=1}^{k_n}\left(F_{nj}(x) - \frac{1}{2} - \frac{1}{2}\mathrm{sgn}(x)\right) = W(x). \tag{4.74}$$

The functional form for the domain of attraction for distributions G characterizing limiting sub-families for the domain of attraction $\mho$ was already seen to be given, in Definition 4.17, by the representation

$$W(dx) = \frac{P}{x^{1+\alpha}}\mathbb{I}_{(0,\infty)}(x)\,dx + \frac{Q}{|x|^{1+\alpha}}\mathbb{I}_{(-\infty,0)}(x)\,dx, \tag{4.75}$$

which can be rewritten in terms of the function for constants C_1 and C_2 according to

$$W(x) = \begin{cases} -C_1 x^{-\alpha}, & \text{if } x > 0, \\ C_2(-x)^{-\alpha}, & \text{if } x < 0, \end{cases} \tag{4.76}$$

for C_1, C_2 non-negative numbers and $\alpha \in (0, 2)$.

Step 2: Consider the restricted setting in which $0 < \alpha < 2$, and make the substitutions of the function for the spectral measure (Lévy spectral measure $W(dx)$) into the log-characteristic function for the Lévy–Khinchin representation resulting from Theorem 4.13 and specified in Definition 4.16 by the representation where we set $b = \sigma^2/2$ to obtain

$$\ln(\Phi_X(\theta)) = \mathrm{i}a\theta - b\theta^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta x \mathbb{I}_{|x|<1})\, W(dx), \tag{4.77}$$

then one may write the Lévy–Khinchin representation slightly differently with regard to a sinusoid truncation function (using the change in the associated drift function) to obtain, under the log transformation, the following representation of the log-characteristic function,

$$\begin{aligned} \ln \mathbb{E}[\exp(\mathrm{i}\theta X)] &= \mathrm{i}a\theta - b\theta^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta \sin x) W(dx) \\ &= \mathrm{i}a\theta - b\theta^2 - C_1 \int_0^\infty (e^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta \sin x) dx^{-\alpha} \\ &\quad - C_2 \int_0^\infty (e^{\mathrm{i}\theta x} - 1 + \mathrm{i}\theta \sin x) dx^{-\alpha} \\ &= \mathrm{i}a\theta - b\theta^2 + C_1 Q_\alpha + C_2 \overline{Q}_\alpha. \end{aligned} \tag{4.78}$$

Note, this is where the C_1 and C_2 in the specification of the Lévy spectral measure of a table distribution come from, see Definition 4.17.

Step 3: Now it will be beneficial to re-represent the function Q_α according to the following function given by

$$\varphi_\alpha(s, p) = p \int_0^\infty (e^{-sx} - e^{-px}) x^{-\alpha} dx, \quad 0 < \alpha < 2, \tag{4.79}$$

in the domain $\mathscr{R}e\{s\}, \mathscr{R}e\{p\} \in \mathbb{R}^+$. In addition, $\varphi_\alpha(s, p)$ is continuous with respect to α in $(0, 2)$ given s, p and continuous on the boundary of the domain on lines $\mathscr{R}e\{s\} = 0$ and $\mathscr{R}e\{p\} = 0$ given α. In addition, the following relationship can be shown relating this function with the Q_α function above,

$$Q_\alpha = \mathscr{R}e\{\varphi_\alpha(\mathrm{i}, p)\}|_{p=-\mathrm{i}\theta}. \tag{4.80}$$

The reason for this re-expression is that the integral defining $\varphi_\alpha(s,p)$ can now be taken, where one considers the case in which $\alpha \neq 1$ and take the limit after first integrating by parts

$$\begin{aligned}\varphi_\alpha(s,p) &= p\int_0^\infty (e^{-sx} - e^{-px})x^{-\alpha}\,dx \\ &= \frac{p}{1-\alpha}\int_0^\infty (se^{-sx} - pe^{-px})x^{1-\alpha}\,dx \\ &= p\frac{\Gamma(2-\alpha)}{1-\alpha}(s^{\alpha-1} - p^{\alpha-1}).\end{aligned} \tag{4.81}$$

Step 4: Next take the limit as $\alpha \to 1$ to obtain

$$\lim_{\alpha\to 1}\varphi_\alpha(s,p) = p\ln(p/s). \tag{4.82}$$

Consequently, Zolotarev (1986) showed that the function Q_α could, therefore, be represented according to

$$\begin{aligned}Q_\alpha &= \mathscr{R}e\{\varphi_\alpha(\mathrm{i},p)\}|_{p=-i\theta} \\ &= \begin{cases} -|\theta|^\alpha\Gamma(1-\alpha)\cos\left(\frac{\pi\alpha}{2}\right) - \mathrm{i}\theta(1-|\theta|^{\alpha-1})\Gamma(1-\alpha)\sin\left(\frac{\pi\alpha}{2}\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha \neq 1, \\ -|\theta|\frac{\pi}{2} - \mathrm{i}\theta\ln|\theta|, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha = 1. \end{cases}\end{aligned} \tag{4.83}$$

Step 5: One then defines the constant d according to

$$d = (C_1 + C_2)\Gamma(2-\alpha)\frac{\sin(\pi(1-\alpha)/2)}{1-\alpha}, \tag{4.84}$$

which would then allow one to write the characteristic function $\Phi_X(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta X)]$ according to

$$|\Phi_X(\theta)| = \exp(-b\theta^2 - d|\theta|^\alpha). \tag{4.85}$$

Step 6: To complete the α-stable characteristic function (Type A parameterisations) derivation, one shows that constants b and d cannot be simultaneously non-zero, see details in (Zolotarev 1986, p. 10–11). This then results in the definition of the characteristic function for a stable random variable of type A, under Zolotarev's notation, given by parameters

$$\begin{aligned}\lambda &= \begin{cases} d, & \text{if } C_1 + C_2 > 0, \\ b, & \text{if } C_1 + C_2 = 0, \end{cases} \\ \beta &= \begin{cases} \frac{C_1 - C_2}{C_1 + C_2}, & \text{if } C_1 + C_2 > 0, \\ 0, & \text{if } C_1 + C_2 = 0, \end{cases} \\ \gamma &= \frac{a + \widetilde{a}}{\lambda},\end{aligned} \tag{4.86}$$

with

$$\widetilde{a} = \begin{cases} (C_2 - C_1)\,\Gamma(1-\alpha)\sin(\pi\alpha/2), & \text{if } \alpha \neq 1, \\ 0, & \text{if } \alpha = 1. \end{cases} \tag{4.87}$$

In the following subsection, we will make explicit several different parameterizations that then follow from this A-type parameterizations. Note, we change the parameter representations relative to those utilized in Zolotarev and other manuscripts so that a common set of parameter labels which are labelled with a sub-index to denoted the 'type' of parameterizations utilized. We believe that this will help to avoid some of the confusions that have abounded in the use of α-stable models in this regard that can occasionally be found in the literature, see discussion in Hall (1981c).

4.4 Popular Parameterizations of the α-Stable Severity Model Characteristic Functions

In this section, we present different aspects of parameterizations of the stable characteristic function found throughout the literature for stable models, as detailed extensively in Zolotarev (1986), Samorodnitsky and Taqqu (1994a) and Nolan (2007).

Definition 4.18 *Denote by $\Phi_X(\theta)$, the characteristic function of a stable univariate distributed random variable $X \stackrel{d}{=} aZ + b$ with $a > 0$, $b \in \mathbb{R}$, $-1 \leq \beta \leq 1$ and $\alpha \in (0, 2]$ and random variable Z has characteristic function given by*

$$\Phi_Z(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta Z)] = \begin{cases} \exp\left(-|\theta|^{\alpha}\left[1 - \mathrm{i}\beta\tan\frac{\pi\alpha}{2}\,(\mathrm{sign}\theta)\right]\right), & \text{if } \alpha \neq 1, \\ \exp\left(-|\theta|\left[1 + \mathrm{i}\beta\frac{2}{\pi}\,(\mathrm{sign}\theta)\ln|\theta|\right]\right), & \text{if } \alpha = 1, \end{cases} \tag{4.88}$$

where

$$\mathrm{sign}\theta = \begin{cases} -1, & \theta < 0, \\ 0, & \theta = 0, \\ 1, & \theta > 0, \end{cases} \tag{4.89}$$

defining $0 \ln 0$ *as* 0. *See further details in Nolan (2007), Samorodnitsky and Taqqu (1994a) and Zolotarev (1986).*

This defines the general parameterizations of a stable random variable; however, the literature contains several alternative parameterizations, each of which has some useful practical properties. In the following, we will present a small subset of some of the more prevalent representations developed and used in the literature, referring the interested reader to advanced discussions in Samorodnitsky and Taqqu (1997) and Nolan (2008).

Remark 4.14 *The following parameterizations will be generally referenced relative to Definition 4.18 where this representation is treated like the form of the standardized univariate α-stable variable which has zero location and unit scale parameter. If the parameters differ from this base representation, we will refer to them explicitly by a subscript for the transformation, for instance,*

$S_\alpha(x;\beta,\gamma_k,\delta_k;k)$ means that parameterization k has a γ and δ parameter which differs under parameterization k relative to the 'standardized' definition provided in Definition 4.18.

4.4.1 UNIVARIATE α-STABLE PARAMETERIZATIONS OF ZOLOTAREV A, M, B, W, C AND E TYPES

Before we present the key parameterizations for the α-stable severity models, we first make the following remark regarding parameterizations.

Remark 4.15 *We note that in this section, we have adopted the standard notation we will utilize through the remainder of the book, replacing the previously used parameters α, β, λ and γ with the more standard notations from Nolan given by α, β, γ and δ. In addition, when required to distinguish the parameterizations, a subscript will be added for the parameterizations considered under Zolotarev's specifications, such as for the A-Type $\alpha_A, \beta_A, \gamma_A$ and δ_A. In addition, it is important to note that the actual understanding of the influence of each parameter on the properties of the resulting stable distribution will be dependent on the parameterizations considered.*

In the foundational book on univariate α-stable models by Zolotarev (1986), there were several parameterizations of the univariate α-stable models developed. These were denoted by the A-type, the M-type and the Q-type. We saw previously how one could derive the characteristic function of the A-type from first principles. In this section, we briefly comment on their representation and properties, starting with the A-type which has characteristic function given in Definition 4.19.

Definition 4.19 (Zolotarev's A-Type Stable Parameterizations) *A random variable X with α-stable distribution, denoted by $X \sim S_\alpha(x;\beta,\gamma_A,\delta_A;A)$ denotes the univariate four-parameter stable distribution family under parameterizations A-type of Zolotarev (1986, p. 9) with characteristic function in the following form,*

$$\Phi_X(\theta) = \mathbb{E}\left[e^{i\theta X}\right] = \begin{cases} \exp\left(i\delta_A\gamma_A\theta - \gamma_A|\theta|^\alpha\left\{1 - i\beta\text{sign}(\theta)\tan\left(\frac{\pi\alpha}{2}\right)\right\}\right) & \alpha \neq 1, \\ \exp\left(i\delta_A\gamma_A\theta - \gamma_A|\theta|\left\{1 + i\beta\text{sign}(\theta)\frac{2}{\pi}\ln|\theta|\right\}\right) & \alpha = 1, \end{cases} \tag{4.90}$$

with stability index $\alpha \in (0,2]$, skewness $\beta \in [-1,1]$, 'rate' $\gamma_A > 0$ and location $\delta_A \in \mathbb{R}$.

It is shown in Cheng and Liu (1997) that the A-type parameterizations has a sharp discontinuity at $\alpha = 1$ for all non-symmetric members of the family, $\beta \neq 0$. This was also noted in Zolotarev (1986, p. 11) when the characteristic function for the A-type parameterizations is derived, it is observed that the characteristic function and, therefore, the distribution function have discontinuities at all points of the form $\alpha = 1$ and $\beta \neq 0$. In addition, for the A-type parameterizations taking the limits as $\alpha \to 1$ with $\beta^* \to \beta$, $\gamma^* \to \gamma$ and $\delta^* \to \delta$ will not only yield an α-stable distribution with parameters $S(x;1,\beta,\gamma,\delta;A)$ but it does not even yield a proper distribution because the measure goes to infinity. For this reason, the Zolotarev (1986) derived the M-type and B-type parameterizations.

EXAMPLE 4.3 Plots of the A-Type Stable Characteristic Function

In this example, we plot the characteristic function of the A-type parameterization for a range of parameter values (Figure 4.2). Each plot shows a range of values of $\theta \in [-10, 10]$, and the radius and phase of the characteristic function are plotted for each value of θ.

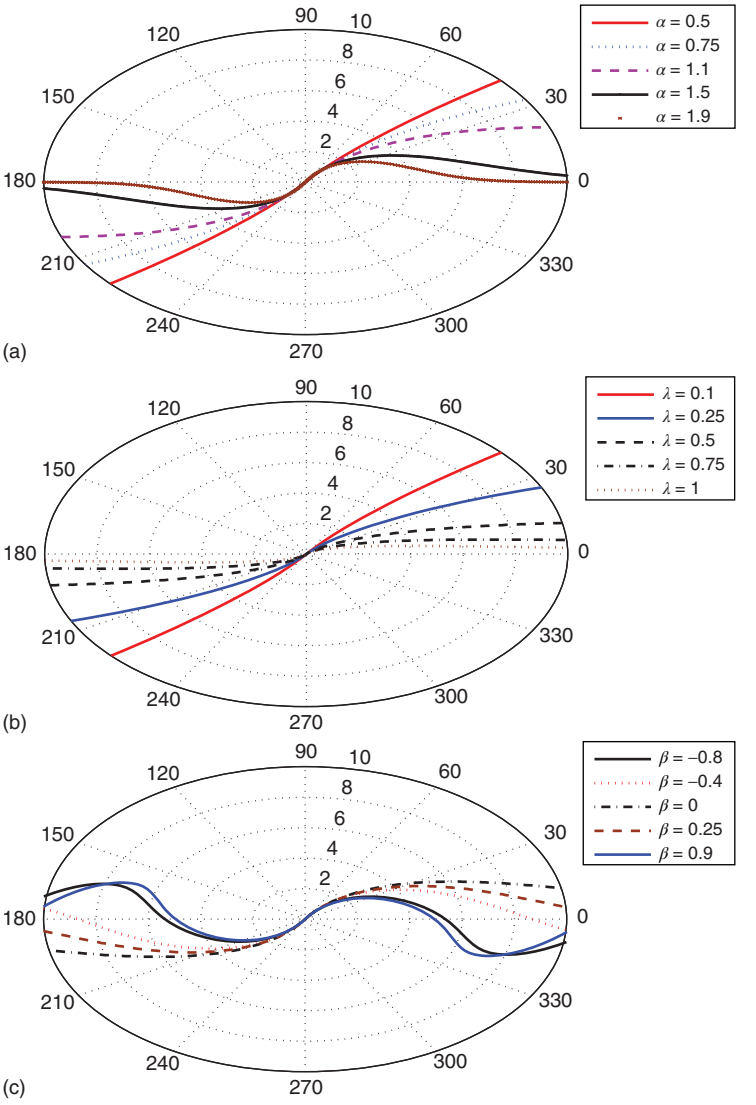

FIGURE 4.2 **(a) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\alpha \in \{0.5, 0.75, 1.1, 1.5, 1.9\}$ with $\beta = 0$, $\gamma = 0.1$ and $\delta = 0$. (b) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\gamma \in \{0.1, 0.25, 0.5, 0.75, 1\}$ with $\alpha = 0.5$, $\beta = 0$ and $\delta = 0$. (c) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\beta \in \{-0.8, -0.4, 0, 0.25, 0.9\}$ with $\alpha = 1.2$, $\gamma = 0.1$ and $\delta = 0$.**
(***See insert for color representation of this figure.***)

The development of the M-type parameterizations in Zolotarev (1986, p. 11) reparameterizes the A-type representation of the characteristic function in such a manner as to ensure that the discontinuity observed in the A-type is removed, thereby obtaining a continuous parameterization in all parameters. This is achieved by the change of variable in the location parameter given by

$$\delta_M = \delta_A - \beta \tan\left(\frac{\pi\alpha}{2}\right), \tag{4.91}$$

which results in the M-type location/scale family. Note that the other parameters are common ($\alpha_A = \alpha_M$, $\beta_A = \beta_M$, $\gamma_A = \gamma_M$) between the A-type and M-type parameterizations and have the same domain over which they are defined. This family is known to have a smooth unimodal density and distribution function. We note that the S0 parameterization of Nolan (2015) is a variant of the M-type parameterization of Zolotarev (1986) with the characteristic function and hence the distribution and density jointly continuous in all four parameters.

It should be noted that the A-type parameterization will experience an asymptotic behaviour for the characteristic function when $\beta \neq 0$ as $\alpha \to 1$ from above or below. In terms of the behaviour of the characteristic function of the A-type parameterization as this asymptote is approached, one will observe increasingly many oscillations as depicted in the following example.

EXAMPLE 4.4 Characteristic Function Asymptotic Boundary $\alpha \to 1$

In this example, we plot the characteristic function of the A-type parameterization for a range of parameter values with $\beta = 0.8$, $\gamma = 0.1$, $\delta = 0$ and $\alpha \in \{0.7, 0.95, 0.99, 1.05, 1.5\}$.

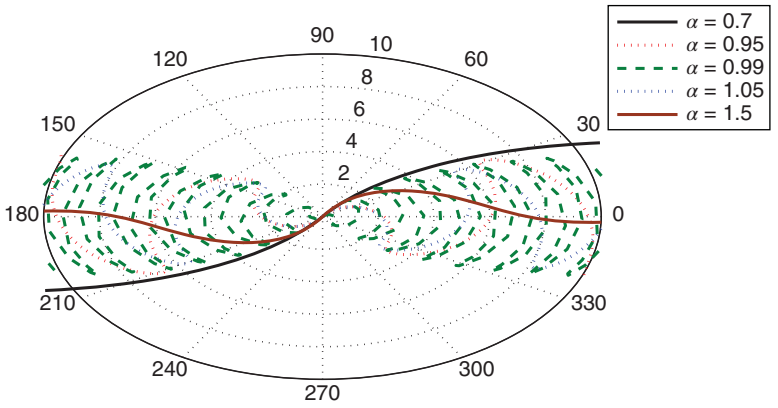

FIGURE 4.3 Example of A-type paramerization for a range of stability index values (See insert for color representation of this figure.)

Then under the B-type parameterization, we see that this asymptotic behaviour around $\alpha \to 1$ is now resolved via the reparameterization. In the following figures, we plot the B-type parameterization with parameters with $\beta = 0.8$, $\gamma = 0.1$, $\delta = 0$ and $\alpha \in \{0.7, 0.95, 0.99, 1.05, 1.5\}$.

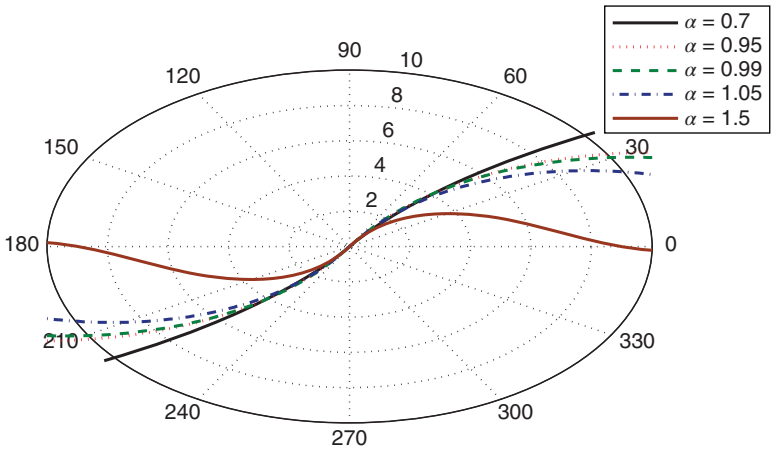

FIGURE 4.4 **Example of B-type paramerization for a range of stability index values (See insert for color representation of this figure.)**

One can overcome this issue with the asymptotic behaviour in the parameter space at $\alpha = 1$ for $\beta \neq 0$ through a reparameterization. The typical choice is to use the B-Type parameterization of Zolotarev (1986), given in Definition 4.20, the problem of non-existence of a proper distribution when taking the limits as $\alpha \downarrow 1$ with $\beta^* \to \beta$, $\gamma^* \to \gamma$ and $\delta^* \to \delta$ is now removed. Consequently, limits will now yield an α-stable distribution with parameters $S(x; 1, \beta, \gamma, \delta; B)$ for $\alpha \geq 1$ and characteristic function given in Equation 4.92.

Definition 4.20 (Zolotarev's B-Type Stable Parameterization) *A random variable X with α-stable distribution, denoted by $X \sim S_\alpha(x; \beta, \gamma_B, \delta_B; B)$, denotes the univariate four-parameter stable distribution family under B-type parameterization of Zolotarev (1986, theorem C.3) with characteristic function in the following form,*

$$\Phi_X(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta X)] = \begin{cases} \exp\left(\gamma_B \left(\mathrm{i}\theta\delta_B - |\theta|^\alpha \exp\left(-\mathrm{i}(\pi/2)\beta K(\alpha)\mathrm{sgn}\theta\right)\right)\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad \alpha \neq 1, \\ \exp\left(\gamma_B(\mathrm{i}\theta\delta_B - |\theta|^\alpha(\pi/2 + \mathrm{i}\beta \ln|\theta|\mathrm{sgn}\theta))\right), \\ \qquad\qquad\qquad\qquad\qquad\qquad \alpha = 1, \end{cases} \tag{4.92}$$

with $K(\alpha) = \alpha - 1 + \mathrm{sgn}(1-\alpha)$ and stability index $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, 'rate' $\gamma_B > 0$ and location $\delta_B \in \mathbb{R}$.

The relationship between the A-type and B-type parameterizations is given in Corollary 4.3, see Zolotarev (1986).

Corollary 4.3 (Relationship Between A-Type and B-Type Parameterizations) *To find the relationship between the A-type and B-type parameterizations, simply equate each side of the characteristic function and rearrange to obtain the following. If $\alpha = 1$, then*

$$\beta_A = \beta_B, \quad \gamma_A = \frac{\pi}{2}\gamma_B, \quad \delta_A = \frac{2}{\pi}\delta_B,$$

and if $\alpha \neq 1$, *then*

$$\beta_A = \cot\left(\frac{\pi}{2}\alpha\right)\tan\left(\frac{\pi}{2}\beta_B K(\alpha)\right),$$
$$\gamma_A = \cos\left(\frac{\pi}{2}\beta_B K(\alpha)\right)\gamma_B,$$
$$\delta_A = \frac{\delta_B}{\cos(2/\pi)\beta_B K(\alpha)}.$$

To complete this section on specifications of the stable models, we also summarize the class of strictly stable models (see Definition 4.7) that are termed by Zolotarev as the W-class and contain two additional parameterizations of interest, the C-type and E-type parameterizations.

Remark 4.16 *It is shown in Zolotarev (1986, I.24) that the definition of strictly stable random variables can be generalized to the case of multiple random variables. As discussed in the Characterization 2, we considered the distribution of* $X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \cdots \stackrel{d}{\neq} \text{const}$ *will be stable iff for any* $n \geq 2$ *there exist numbers* $B_n > 0$ *and* $A_n \in \mathbb{R}$ *such that*

$$X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \cdots \stackrel{d}{=} X_n \stackrel{d}{=} B_n X_1 + A_n. \tag{4.93}$$

It was shown in Zolotarev (1986) that this criterion is both necessary and sufficient to characterize the family of stable distributions in general and in particular the family of strictly stable distributions.

In Proposition 4.2, a criterion for a non-degenerate severity distribution to belong to the class of strictly stable distributions is provided.

Proposition 4.2 (Criterion for Strict Stability: Defining the W-Class) *In Kagan* et al. *(1973), the following conditions are provided for membership of a non-degenerate severity distribution to the class of strictly stable distributions, terms the W-class. Let* $c_j = \exp(-d_j), j = 1, \ldots, k, k \geq 2$ *be constants that satisfy the following conditions:*

1. $c_1^2 + \cdots + c_k^2 \leq 1$,
2. *there are two numbers* d_i *and* d_j *from the corresponding set* $\{d_1, d_2, \ldots, d_k\}$ *with an irrational ratio.*

Then, given these constants, the distribution of the random variables

$$X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \cdots \stackrel{d}{=} X_k \stackrel{d}{\neq} \text{const}$$

is in the strictly stable W-class as long as the linear combination given by $\sum_{i=1}^{k} c_i X_i$ *is equal in distribution to* X_1.

It was then shown in Zolotarev (1986) that for such strictly stable members, one can consider two additional parameterizations denoted by C- and E-types as detailed in Definition 4.21.

Definition 4.21 (Zolotarev's C-Type and E-Type Strictly Stable Parameterizations) *The following are two parameterizations for the strictly stable severity distribution log-characteristic functions denoted by prefixes C and E:*

1. *The strictly stable distribution* $S_\alpha(x;\beta_C,\gamma_C;C)$ *has characteristic functions*

$$\Phi_X(\theta)=\mathbb{E}[\exp(\mathrm{i}\theta X)]=\exp\Big(-\gamma_C|\theta|^\alpha\exp\Big(-\mathrm{i}\frac{\pi}{2}\beta_C\alpha\mathrm{sgn}\theta\Big)\Big) \tag{4.94}$$

with parameters $\alpha\in(0,2]$, $|\beta_C|\le\min(1,2/\alpha-1)$ *and* $\gamma_C>0$.

2. *The strictly stable distribution* $S_\alpha(x;\beta_E,\gamma_E;E)$ *has characteristic functions*

$$\begin{aligned}\Phi_X(\theta)&=\mathbb{E}[\exp(\mathrm{i}\theta X)]\\&=\exp\left(-\exp\Big(\alpha^{-\frac{1}{2}}\left(\ln|\theta|+\gamma_E-\mathrm{i}\left(\frac{\pi}{2}\right)\beta_E\mathrm{sgn}\theta\right)+\mathbb{C}\left(\alpha^{-\frac{1}{2}}-1\right)\Big)\right)\end{aligned} \tag{4.95}$$

with Euler constant $\mathbb{C}=\lim_{n\to\infty}\left(\sum_{k=1}^n k^{-1}-\ln(n)\right)=0.577\ldots$ *and parameters* $\alpha\in[0.25,\infty)$, $|\beta_E|\le\min(1,2\sqrt{\alpha}-1)$ *and* $|\gamma_E|<\infty$.

Remark 4.17 (Relationship Between B-Type, C-Type and E-Type Parameterizations) *When* $\alpha=1$ *and* $|\beta|=1$, *one obtains the degenerate distribution at the point* $\beta\gamma$. *In addition, excluding this point parameter set* $\alpha=1$ *and* $|\beta|=1$, *the C-type parameterization is related back to the B-type as follows:*

$$\begin{aligned}\alpha_C&=\alpha_B,\\ \beta_C&=\begin{cases}\beta_B K(\alpha)/\alpha, & \text{if } \alpha\neq1,\\ \frac{2}{\pi}\arctan\left(2\delta_B/\pi\right), & \text{if } \alpha=1,\end{cases}\\ \gamma_C&=\begin{cases}\gamma_B, & \text{if } \alpha\neq1,\\ \gamma_B\left(\pi^2/4+\delta_B^2\right), & \text{if } \alpha=1.\end{cases}\end{aligned} \tag{4.96}$$

Remark 4.18 *One can connect the E-type parameterization to the C-type parameterization as follows:*

$$\begin{aligned}\alpha_E&=\alpha_C^{-2},\\ \beta_E&=\beta_C,\\ \gamma_E&=\frac{1}{\alpha_C}\ln\gamma_C+\mathbb{C}\left(\frac{1}{\alpha_C}-1\right).\end{aligned} \tag{4.97}$$

In each of the parameterizations, it is noted in (Zolotarev 1986, property 2.1) that one has translation and scale invariance of the stable law, as stated in Lemma 4.3.

Lemma 4.3 (Scaling and Translation Distributional Invariance for Stable Laws) *If* $Z\sim S(\alpha,\beta,\gamma_1,\delta_1;\cdot)$, *then for any* $a\neq0,b\in\mathbb{R}$, *the transformation* $X=aZ+b$ *is a scaled version of the* α*-stable distribution such that*

$$X\stackrel{d}{=}aZ+\gamma_2 b, \tag{4.98}$$

where $X\sim S(\alpha,\beta,\gamma_2,\delta_2;\cdot)$.

An example for the case of the A-type parameterization is provided, where one obtains

$$\begin{aligned} a &= \left(\frac{\gamma_2}{\gamma_1}\right)^{1/\alpha}, \\ b &= \begin{cases} \delta_2 - \delta_1(\gamma_2/\gamma_1)^{1/(\alpha-1)}, & \alpha \neq 1, \\ \delta_2 - \delta_1 + \frac{2}{\pi}\beta \ln(\gamma_2/\gamma_1), & \alpha = 1. \end{cases} \end{aligned} \tag{4.99}$$

Next we present two additional approaches adopted in the literature that are popular parameterizations for the stable family, derived in Nolan (2015).

4.4.2 UNIVARIATE α-STABLE PARAMETERIZATIONS OF NOLAN S0 AND S1

Consider the generic parameterization of the α-stable random variables characteristic function given in Definition 4.19 which is Zolotarev's A-type parameterization with $\gamma_A = 1$ and $\delta_A = 0$ which we denote here again by the reference 'standardized' stable random variable by Z which has characteristic function given by

$$\Phi_Z(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta Z)] = \begin{cases} \exp(-|\theta|^\alpha \left[1 - \mathrm{i}\beta \tan(\pi\alpha/2)\,(\mathrm{sign}\theta)\right]), & \text{if } \alpha \neq 1, \\ \exp\left(-|\theta| \left[1 + \mathrm{i}\beta\frac{2}{\pi}\,(\mathrm{sign}\theta)\ln|\theta|\right]\right), & \text{if } \alpha = 1. \end{cases} \tag{4.100}$$

From this definition, two parameterizations were proposed by Nolan (2015) and are known as the $S0$ and $S1$ parameterizations detailed in Definition 4.22.

Definition 4.22 (Nolan's $S0$ and $S1$-Type Stable Parameterizations) *The following are two popular parameterizations presented in Nolan (2015) that are denoted by $S0$ and $S1$ prefixes.*

- $X \sim S\,(\alpha, \beta, \gamma_{S0}, \delta_{S0}; S0)$ *iff*

$$X \overset{d}{=} \begin{cases} \gamma_{S0}\,(Z - \beta\tan(\pi\alpha/2)) + \delta_{S0}, & \alpha \neq 1, \\ \gamma_{S0}Z + \delta_{S0}, & \alpha = 1, \end{cases} \tag{4.101}$$

  where Z has characteristic function given in Definition 4.18. Under this parameterization, α and β determine the shape of the distribution, γ_{S0} the scale and δ_{S0} the location. The resulting characteristic function for the $S0$ parameterization has the form

$$\mathbb{E}[\exp(\mathrm{i}\theta X)] = \begin{cases} \exp\{-\gamma_{S0}^\alpha|\theta|^\alpha(1 + \mathrm{i}\beta\mathrm{sign}(\theta)\tan(\frac{\pi\alpha}{2})(|\gamma_{S0}\theta|^{1-\alpha} - 1)) + \mathrm{i}\delta_{S0}\theta\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha \neq 1, \\ \exp\{-\gamma_{S0}|\theta|(1 + \mathrm{i}\beta\frac{2}{\pi}\mathrm{sign}(\theta)\ln(\gamma_{S0}|\theta|)) + \mathrm{i}\delta_{S0}\theta\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \alpha = 1. \end{cases} \tag{4.102}$$

- $X \sim S(\alpha, \beta, \gamma_{S1}, \delta_{S1}; S1)$ *iff*

$$X \overset{d}{=} \begin{cases} \gamma_{S1}Z + \delta_{S1}, & \alpha \neq 1, \\ \gamma_{S1}Z + \left(\delta_{S1} + \beta\frac{2}{\pi}\gamma_{S1}\ln\gamma_{S1}\right), & \alpha = 1, \end{cases} \tag{4.103}$$

where Z has characteristic function given in Definition 4.18. The resulting characteristic function for the $S1$ parameterization has the form

$$\mathbb{E}[\exp(\mathrm{i}\theta X)] = \begin{cases} \exp\{-\gamma_{S1}^{\alpha}|\theta|^{\alpha}\left[1 - \mathrm{i}\beta \mathrm{sign}(\theta)\tan(\frac{\pi\alpha}{2})\right] + \mathrm{i}\delta_{S1}\theta\}, \text{ if } \alpha \neq 1, \\ \exp\{-\gamma_{S1}|\theta|\left[1 + \mathrm{i}\beta(\frac{2}{\pi})\mathrm{sign}(\theta)\ln|\theta|\right] + \mathrm{i}\delta_{S1}\theta\}, \text{ if } \alpha = 1. \end{cases} \tag{4.104}$$

As discussed in Nolan (2015), the parameterization $S(\alpha, \beta, \gamma_{S0}, \delta_{S0}; S0)$ and $S(\alpha, \beta, \gamma_{S1}, \delta_{S1}; S1)$ are identical in the symmetric case in which $\beta = 0$. Furthermore, it is worth observing that $S(2, 0, \gamma_{S0}, \delta_{S0}; S0) = N(\delta_{S0}, 2\gamma_{S0}^2)$, and in addition, under the $k = 0$ parameterization, one obtains the property of standardization given by $(X - \delta_{S0})/\gamma_{S0} \sim S(\alpha, \beta, 1, 0; S0)$. Finally, these parameterizations for $S0$ and $S1$ share common α, β and γ values and the mapping between the two parameterizations can be obtained by

$$\begin{aligned} \delta_{S0} &= \begin{cases} \delta_{S1} + \beta\gamma\tan(\pi\alpha/2), & \alpha \neq 1, \\ \delta_{S1} + \beta\frac{2}{\pi}\gamma\ln\gamma, & \alpha = 1. \end{cases} \\ \delta_{S1} &= \begin{cases} \delta_{S0} - \beta\gamma\tan(\pi\alpha/2), & \alpha \neq 1, \\ \delta_{S0} - \beta\frac{2}{\pi}\gamma\ln\gamma, & \alpha = 1. \end{cases} \end{aligned} \tag{4.105}$$

In the remainder of this section, we summarize a few basic properties of the stable family parameterizations presented in Definition 4.22. In particular, we relate these properties to relevant application scenarios in the OpRisk severity modeling situation under an LDA framework.

We first present in Lemma 4.4 the support of a random variable $X \sim S(\alpha, \beta, \gamma_k, \delta_k; k)$, $k \in \{S0, S1\}$ as a function of the distributional parameters, see Nolan (2015).

Lemma 4.4 *Denote the sample space S_X or support of the distribution of a univariate random variable $X \sim S(\alpha, \beta, \gamma_k, \delta_k; k)$, $k \in \{S0, S1\}$ as follows:*

$$S_X = \mathrm{Supp}(S(\alpha, \beta, \gamma_{S0}, \delta_{S0}; S0)) = \begin{cases} \left[\delta_{S0} - \gamma_{S0}\tan\frac{\pi\alpha}{2}, \infty\right), & \alpha < 1 \text{ and } \beta = 1, \\ \left(-\infty, \delta_{S0} + \gamma_{S0}\tan\frac{\pi\alpha}{2}\right], & \alpha < 1 \text{ and } \beta = -1, \\ (-\infty, \infty), & \text{otherwise.} \end{cases}$$

$$S_X = \mathrm{Supp}(S(\alpha, \beta, \gamma_{S1}, \delta_{S1}; S1)) = \begin{cases} [\delta_{S1}, \infty), & \alpha < 1 \text{ and } \beta = 1, \\ (-\infty, \delta_{S1}], & \alpha < 1 \text{ and } \beta = -1, \\ (-\infty, \infty), & \text{otherwise.} \end{cases}$$

It is worth noting the following points about these two parameterizations; it is recommended in Nolan (2015) that for numerical work, the $S0$ parameterization is utilized because it has the simplest form and the characteristic function is continuous in all parameters avoiding complications with inference under this parameterization. Furthermore, the random number generator for a univariate stable distribution developed by Chambers *et al.* (1976) will produce draws according to the $S0$ parameterization. Often, in practice, for derivation of properties of the α-stable model, the $S1$ parameterization will be adopted because of its algebraic properties.

The following remark regarding the $S1$ parameterization should be considered when working with this parameterization of the model in practice.

Remark 4.19 *Under the $S0$ parameterization of the α-stable model, the mode of the distribution will be unbounded for any α in the neighborhood of 1, i.e. $\alpha \in$ n.e.(1), with the following asymptotic limits for the behaviour of the mode in the neighbourhood of $\alpha = 1$:*

$$\text{Mode}(X) = \begin{cases} \lim\limits_{\alpha \uparrow 1} \arg\ \max_x S(x; \alpha, \beta, \gamma_{S1}, \delta_{S1}; S1) \to \infty, & \text{if } \beta > 0, \\ \lim\limits_{\alpha \downarrow 1} \arg\ \max_x S(x; \alpha, \beta, \gamma_{S1}, \delta_{S1}; S1) \to -\infty, & \text{if } \beta < 0. \end{cases} \tag{4.106}$$

Furthermore, the $S1$ parameterization will not be continuous in all parameters and nor will it provide a scale-location family of distributions.

In addition, under the $S0$ parameterization that we consider in the remainder of this section, the mode, the percentiles and the Paretian tail behaviour all vary in a continuous manner as α and β vary. Having defined these parameterizations of the α-stable, one can also show that linear transformations of α-stable random variables preserves the distributional property of stability, as shown in Lemma 4.5.

Lemma 4.5 (Scaling and Translation Distributional Invariance for $S0$) *If $Y \sim S(\alpha, \beta, \gamma_{S0}, \delta_{S0}; S0)$, then for any $a \neq 0, b \in \mathbb{R}$, the transformation $Z = aY + b$ is a scaled version of the α-stable distribution. That is, $Z \sim S_\alpha(\text{sign}(a)\beta, |a|\gamma_{S0}, a\delta_{S0} + b; S0)$. In addition, the characteristic function, densities and distribution functions are jointly continuous in all four parameters $(\alpha, \beta, \gamma_{S0}, \delta_{S0})$ and in x. These results follow from Samorodnitsky and Taqqu (1994b) and Nolan (2013, proposition 1.16).*

Before progressing to the density, distribution and quantile function representations, we note the following interesting property of α-stable severity models. The α-stable distribution can be represented according to the following identities given in Proposition 4.3, see Zolotarev (1986).

Proposition 4.3 (α-Stable Severity Model Representations) *Given a loss random variable under Nolan's $S1$ parameterization $X \sim S(\alpha, \beta, \gamma_{S1}, \delta_{S1} = 0; S1)$, there exist two i.i.d. perfectly skewed α-stable random variables Y_1 and Y_2 with common distribution $Y_1 = Y_2 = S(\alpha, \beta = 1, \gamma_{S1}, \delta_{S1} = 0; S1)$ such that one has the representation*

$$X \stackrel{d}{=} \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_1 - \left(\frac{1+\beta}{2}\right)^{1/\alpha} Y_2. \tag{4.107}$$

Remark 4.20 *This result is interesting as it shows the importance of the role of the perfectly skewed boundary cases that they can act as building blocks for all other stable random variables and they will also be of particular significance in OpRisk, especially for cases with $\alpha < 1$, where the support of such distributions is strictly positive.*

In the following sections, we describe several relevant properties of α-stable models that make them amenable to consideration as OpRisk severity models for low frequency but high

consequence loss process modeling. This includes the different representations of their density and their properties with respect to linear combinations of losses distributed according to such a severity model. This will be of direct relevance to understanding representations of random sums in an LDA model. The practical utilization of such models requires three key features:

- the ability to evaluate the severity distribution pointwise;
- the ability to simulate from the severity distribution; and
- the ability to perform parameter inference under such a severity model given data.

4.5 Density Representations of α-Stable Severity Models

One of the difficulties faced by practitioners in utilizing α-stable models in practical settings is to understand how best to represent and evaluate the density and distribution functions for these models pointwise. There exist few closed-form simple expressions for such an evaluation; it is, therefore, the intention of this section to clarify exactly what can be achieved in this regard and the different approaches one can adopt in practice. We will start from the perspective that given the characteristic function, we have a unique specification of the distribution function of the severity model, as specified in the uniqueness theorem in Theorem 4.4. As this uniqueness result allows us to recognize that there is a one-to-one correspondence between a cumulative distribution function for a loss random variable and the corresponding characteristic function of the loss random variable, then it is always possible to find one of these functions if we know the other one.

Remark 4.21 *The ability to accurately and efficiently evaluate the density function pointwise for the α-stable distributions arises in many practical contexts when using such models in OpRisk settings, such as evaluation of convolutions, parameter estimation via the likelihood and risk measure evaluation. Hence, we discuss in the following several approaches one may adopt to perform density estimation for the α-stable family.*

In this section, we focus on how to apply an inversion theorem to go from the characteristic function $\Phi_X(\theta)$ of a random variable X to its distribution $F(x)$ or density $f(x)$. This uses the Fourier inversion theorem given in Theorem 4.16.

Theorem 4.16 (Fourier Inversion Theorem) *Consider a random variable X with distribution function $F(x)$, density function $f(x)$ and characteristic function $\Phi_X(\theta)$ such that $f(x)$ and Φ_X are Lebesgue integrable functions. Then given the characteristic function*

$$\Phi_X(\theta) = \int_{-\infty}^{\infty} \exp(\mathrm{i}\theta x) f(x)\, dx = \mathbb{E}[\exp(\mathrm{i}\theta X)], \tag{4.108}$$

one can obtain, via the inverse Fourier transform of the characteristic function, the density

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}\theta x) \Phi_X(\theta)\, d\theta. \tag{4.109}$$

This inversion theorem results in the following relationships stated previously:

1. If the characteristic function $\Phi_X(\theta)$ is integrable, then the distribution $F(x)$ is absolutely continuous and as a result the loss random variable X has a pdf given by the solution to the integral
$$f(x) = \frac{d}{dx}F(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathrm{i}\theta x)\Phi_X(\theta)\, d\theta. \tag{4.110}$$
2. If the characteristic function $\Phi_X(\theta)$ is associated to the distribution $F(x)$ which is continuous at points x_0 and x_1 where $x_0 < x_1$, then the probability $\mathbb{P}\mathrm{r}[x_0 < X < x_1]$ is given by
$$F(x_1) - F(x_0) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T} \frac{\exp(-\mathrm{i}\theta x_0) - \exp(-\mathrm{i}\theta x_1)}{\mathrm{i}\theta}\Phi_X(\theta)\, d\theta. \tag{4.111}$$
3. If a point x_0 is an atom in the support of the distribution $F(x)$, that is, a point of discontinuity, then one has
$$F(x_0) - F(x_0 - 0) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} \exp(-\mathrm{i}\theta x_0)\Phi_X(\theta)\, d\theta. \tag{4.112}$$

In general, these results are stated generally by the theorem of Gil-Palaez given below, see Gil-Pelaez (1951) and Wendel (1961). It tells one that if $\Phi_X(\theta)$ is the characteristic function of a distribution function $F(x) = \frac{1}{2}\{F(x-0) + F(x+0)\}$, then one can obtain an expression for the inverse of the characteristic function to get the distribution even if the integral required is improper at either or both of the end points for the terminals (as indicated by the arrows).

Theorem 4.17 (Gil–Pelaez Inversion Integral) *If X is a univariate loss random variable and x is a continuity point in the support of the distribution function $F(x)$, then the distribution is given by*
$$F(x) = \mathbb{P}\mathrm{r}[X \le x] = \frac{1}{2} - \frac{1}{\pi}\int_{\to 0}^{\to\infty} \frac{1}{\theta}\mathcal{I}\mathrm{m}\{\exp(\mathrm{i}\theta x)\Phi_X(\theta)\}d\theta. \tag{4.113}$$

In the following subsections, we will discuss in some detail different methods one may adopt to both represent the stable distribution functions under different parameterizations and then also to solve these integral representations numerically. Examples of numerical procedures for such integrations abound in the literature, and we provide a brief discussion on a few, see, for instance, Davies (1973), Waller *et al.* (1995), Abate and Whitt (1992), Bohman (1975) and Shephard (1991). In terms of specific examples for stable distributions, there is a good treatment in Nolan (1997). There are also numerous papers on how to calculate bounds for the associated numerical errors such as Hughett (1998).

4.5.1 BASICS OF MOVING FROM A CHARACTERISTIC FUNCTION TO A DISTRIBUTION OR DENSITY

It is clear by now that the specification of the α-stable severity distributions are given according to the characteristic function of the stable law, under one of the many possible parameterizations that have been discussed. It has typically been the focus in the literature to work with the form of Nolan's $S1$ or $S0$ parameterizations. Then, with respect to obtaining the integral representations, there have been three main categories of approximations considered:

- Direct numerical integrations (DuMouchel 1975 and DuMouchel 1971). A procedure for approximating the distributions function via Bergstroms series expansion and Zolotarev's series expansions is developed. In Holt and Crow (1973), a numerical inversion to obtain the pdf is considered; then in Nolan (1999), an approach is developed to evaluate the M-type representation of Zolotarev.
- Fast Fourier Transforms (FFT) and Tail Series Approximations (Mittnik *et al.* 1999a). Consider density estimation via the FFT solution to the integral inversion—note that this approach has upper and lower truncations as well as uniform grid point spacings, but can be highly computationally efficient.
- Methods of Multiple Quadratures (Belov 2005). Considers an approach known as the method of two quadratures.

It is clear that there have been several approaches considered to tackle the density estimation and distribution function estimation. In the following, we consider highlighting aspects of different approaches.

Before proceeding, we will simplify the following analysis by reducing the problem to a simpler set of considerations, where we may without loss of generality remove the need to consider the parameters γ and δ. To achieve this, it is relevant to note the following property of α-stable severity distribution models given in Proposition 4.4 which will allow us to simplify some density representations in this section.

Proposition 4.4 *Given two stable random variables,* $X_1 \sim S_\alpha(x;\beta,\gamma_1,\delta_1;A)$ *and* $X_2 \sim S_\alpha(x;\beta,\gamma_2,\delta_2;A)$*, with common* α *and* β *parameters, one can specify two real numbers* $a>0$ *and* $b\in\mathbb{R}$ *such that the following distributional relationship is satisfied:*

$$X_1 \stackrel{d}{=} aX_2 + \delta_1 b, \tag{4.114}$$

with

$$a = \left(\frac{\delta_1}{\delta_2}\right)^{1/\alpha},$$

$$b = \begin{cases} \gamma_1 - \gamma_2\left(\frac{\delta_1}{\delta_2}\right)^{-1+1/\alpha}, & \text{if } \alpha \neq 1, \\ \gamma_1 - \gamma_2\frac{2}{\pi}\beta\ln\left(\frac{\delta_1}{\delta_2}\right), & \text{if } \alpha = 1. \end{cases}$$

Note that one can also find equivalent representations for other parameterizations such as the result presented in Nolan (2008, propositions 1.16 and 1.17), which show equivalent results under the $S0$ and $S1$ parameterization as detailed in Proposition 4.5.

Proposition 4.5 *Given two stable random variables,* $X_1 \sim S_\alpha(x;\beta_1,\gamma_1,\delta_1;S0)$ *and* $X_2 \sim S_\alpha(x;\beta_2,\gamma_2,\delta_2;S1)$*, one can show that for two real numbers* $a\in\mathbb{R}\backslash\{0\}$ *and* $b\in\mathbb{R}$*, the following distributional relationship is satisfied:*

1. *in the case of the* $S0$ *parameterization, one has the following scale and location invariance*

$$aX_1 + b \sim S_\alpha(\text{sgn}(a)\ \beta_1, |a|\gamma_1, a\delta_1 + b; S0), \tag{4.115}$$

2. *in the case of the* $S1$ *parameterization, one has the following scale and location invariance*

$$aX_2 + b \sim \begin{cases} S_\alpha(\text{sgn}(a)\ \beta_2, |a|\gamma_2, a\delta_2 + b; S1), & \text{if } \alpha \neq 1, \\ S_1(\text{sgn}(a)\ \beta_2, |a|\gamma_2, a\delta_2 + b - \frac{2}{\pi}\beta_2 a \ln|a|; S1), & \text{if } \alpha = 1. \end{cases} \tag{4.116}$$

Given these properties of stable severity models, it, therefore, suffices to consider in this section the case of the 'standardized' α-stable severity distributions in which without loss of generality one considers $\gamma = 1$ and $\delta = 0$. In addition, it will be evident that when working under the B-type parameterization of Zolotarev, one has the following properties for the density, distribution function and characteristic function

$$\begin{aligned} f(-x;\alpha,\beta;B) &= f(x;\alpha,-\beta;B), \\ F(-x;\alpha,\beta;B) &= 1 - F(x;\alpha,-\beta;B), \\ \Phi(-\theta;\alpha,\beta;B) &= \Phi(\theta;\alpha,-\beta;B), \end{aligned} \tag{4.117}$$

which means that one can work purely with the positive axis for $x \geq 0$ and the skewness parameter $\beta \geq 0$ when performing the integration for the evaluation of the density. Proceeding from this assumption, we can then present the solutions for the evaluation of the density of an α-stable severity distribution according to an integral representation via Fourier integrals as follows in Equation 4.120, as long as the selected form of the characteristic function is integrable.

$$\begin{aligned} f(x;\alpha,\beta;\cdot) &= \frac{1}{2\pi}\int \exp(-\mathrm{i}\theta x)\Phi_X(\theta;\alpha,\beta;\cdot)d\theta \\ &= \frac{1}{\pi}\mathscr{R}\mathrm{e}\int_0^\infty \exp(\mathrm{i}\theta x)\Phi_X(\theta;\alpha,-\beta;\cdot)d\theta. \end{aligned} \tag{4.118}$$

From this point, some work on finding appropriate branch cuts of the complex plane to consider integrating along to solve this integral inversion are considered in detail in Zolotarev (1986). In particular, for the B-type integral representation, the results obtained are provided in Zolotarev (1986, Equations 2.2.1a and 2.2.1b), and for the A-type integral representations, the results obtained are provided in (Zolotarev 1986, theorem 2.2.2). However, as noted in Zolotarev, these integral representations are difficult to work with as they provide oscillatory behaviour in the integrand.

- We will discuss an example of the B-type in Density Approximation Approach 1, although we will deal with the oscillation through transformations and weighting, to find an efficient integral approximation given by the Clenshaw–Curtis discrete cosine transform (DCT) quadrature rule.
- Zolotarev also provides a transformation of the oscillatory integral arguments obtained by the A-type and B-type representations which removes the oscillatory behaviour. This form is considered in Density Approximation approaches 3 and 4.

Before getting into particular details, we note that the starting point of all these methods is, therefore, to substitute each of the expressions obtained for the characteristic function under each of the parameterizations into the integral expression in Equation 4.118 in order to find explicit integral representations. This opens up to the following possibilities:

1. For each individual parameterization, find a suitable integral representation that can be solved either analytically or numerically (avoiding issues such as oscillation in the integrand via change of variables or transformations). This was the approach of Zolotarev for the A-type and B-type parameterizations which led to a universal representation given in Zolotarev (1986, theorem 2.2.3) that is considered in Density Approximation approaches 3 and 4.
2. The second alternative is to work with single parameterizations integral representation for the density, such as the B-type considered later and to always apply transformations on the parameters to move from a different parameterization to the chosen parameterization (i.e. B-type). This could involve moving between transformations of the characteristic functions explicitly or between the equivalent transformations of the resulting density. For instance, in Zolotarev (1986, Equations 2.2.16 and 2.2.17), it was shown that if one has the density under the B-type solved for particular points, then one can trivially obtain the density for the M-type parameterizations at particular transformed points, without recalculation, according to the relationships between the actual density values given by

$$f\Big(x+\beta\tan\Big(\frac{\pi}{2}\alpha\Big)\,|x|^{1-\alpha};\alpha,\beta;A\Big)=f\left(x+\beta\tan\left(\frac{\pi}{2}\alpha\right)\left(|x|^{1-\alpha}-1\right);\alpha,\beta;M\right). \tag{4.119}$$

Hence, we first consider the second approach and work with the type-B Zolatarev characteristic function parameterization and the associated integral representation obtained in Zolotarev (1986, Equations 2.2.1a and 2.2.1b) to illustrate how one may work with this under modern numerical approaches.

Definition 4.23 (Zolotarev's B-Type Integral Representations) *Consider* $X\sim S_{\alpha}(\alpha,\beta,1,0;B)$*, then the density can be obtained through the solution to the integral*

$$f\left(x;\alpha,\beta;B\right)=\begin{cases}\frac{1}{2\pi}\mathscr{R}\mathrm{e}\int_0^{\infty}\exp\big(-\mathrm{i}\theta x-\theta^{\alpha}\exp\big(-\mathrm{i}\frac{\pi}{2}\beta K(\alpha)\big)\big)\,d\theta, & \text{if } \alpha\neq 1,\\ \frac{1}{2\pi}\mathscr{R}\mathrm{e}\int_0^{\infty}\exp\big(-\mathrm{i}\theta x-\frac{\pi}{2}\theta-\mathrm{i}\beta\theta\ln\theta\big)\,d\theta, & \text{if } \alpha=1,\end{cases}$$

with $K(\alpha)=\alpha-1+\mathrm{sgn}(1-\alpha)$.

Remark 4.22 *In working with these integrals, we note that because the integrands are simple functions and univariate, one could adopt ones favourite numerical integration technique such as a Gauss–Kronrod quadrature rule and a Clenshaw–Curtis quadrature rule or even a simple Simpsons rule to approximate the integrals over a finite domain. However, it should be noted that this will be complicated for the expression of the integrals above which were naively obtained through Fourier integration of the characteristic function. The reason for this is illustrated in Example 4.5 where it is shown that such integrand functions oscillate by changing sign an infinite number of times. One must, therefore, handle this feature carefully when adopting a quadrature rule to ensure accurate evaluation.*

EXAMPLE 4.5 Characteristic Function Asymptotic Boundary $\alpha \to 1$

For example, if one considers the real and imaginary components of this integrand, given, respectively, for a fixed value of x by

$$\begin{aligned}&\exp\left(-\mathrm{i}\theta x - \theta^{\alpha}\exp\left(-\mathrm{i}\frac{\pi}{2}\beta K(\alpha)\right)\right)\\&\quad= \exp\left\{-\theta^{\alpha}\cos\left(-\frac{\pi}{2}\beta K(\alpha)\right)\right\}\left\{\cos\left[\theta x + \theta^{\alpha}\sin\left(-\frac{\pi}{2}\beta K(\alpha)\right)\right]\right.\\&\quad\left.-\mathrm{i}\sin\left[\theta x + \theta^{\alpha}\sin\left(-\frac{\pi}{2}\beta K(\alpha)\right)\right]\right\},\end{aligned} \tag{4.120}$$

which when one plots the real component, one sees the following oscillatory behaviours for a range of α and β values (Figure 4.5). It is evident that the behaviour of the oscillation is a function of the tail index value α, with a clearly stronger and more sustained oscillation amplitude occurring as α decreases. This behaviour is less affected by the skewness parameter β.

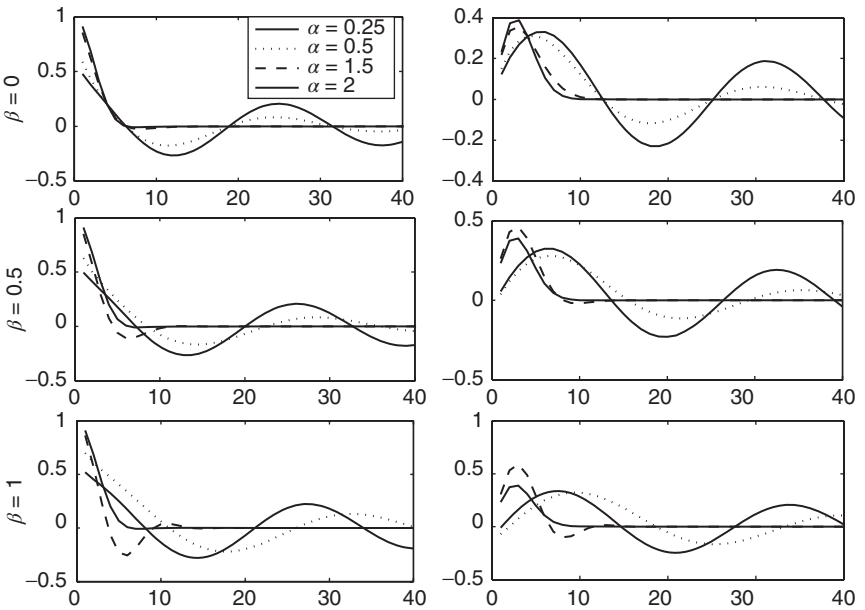

FIGURE 4.5 **(a) displays the real integrand function and (b) displays the imaginary component of the integrand. The three sub-plots on each panel demonstrate a range of values for the symmetry parameter $\beta \in \{0, 0.5, 1\}$ and the four plots in each sub-plot demonstrate the function for a range of tail index $\alpha \in \{0.25, 0.5, 1.5, 2\}$ values. Each plot is for $x = 1$ and is displayed as a function of the argument of the integration θ.**

Approximating these integrals will have three challenges:

- *Challenge 1.* The first will be controlling the discretization error present from the selected quadrature rule;

- *Challenge 2.* The second will be dealing with the truncation error associated with an upper bound on the number of quadrature points; and
- *Challenge 3.* The third will be handling the oscillatory nature of the integrand.

One way to address this oscillatory nature when approximating these intractable integrals with a numerical integration would be to utilize a weighted quadrature rule such as illustrated in the following example of a solution obtained under a Clenshaw–Curtis *weighted* quadrature rule. Other approaches that adopt more sophisticated solutions are provided in Iserles *et al.* (2006), Asheim *et al.* (2012) and Kim *et al.* (2003). To tackle the truncation error, one can make a correction of the form discussed in Luo and Shevchenko (2011) and detailed in brief in next section.

EXAMPLE 4.6

If one considers $X \sim S_{\alpha}(\alpha, \beta, 1, 0; B)$, then the density can be obtained through the solution to the real part of the integral of the complex-valued function given, for example, in the case in which $\alpha \neq 1$ by

$$\begin{aligned} f(x;\alpha,\beta;B) &= \frac{1}{2\pi}\mathscr{R}\mathrm{e}\int_0^{\infty}\exp\Big(-\mathrm{i}\theta x-\theta^{\alpha}\exp\Big(-\mathrm{i}\frac{\pi}{2}\beta K(\alpha)\Big)\Big)\,d\theta \\ &= \frac{1}{2\pi}\int_0^{\infty}\exp\Big\{-\theta^{\alpha}\cos\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big\} \\ &\quad\times\cos\Big[\theta x+\theta^{\alpha}\sin\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big]\,d\theta. \end{aligned} \tag{4.121}$$

The methods presented in the following sections will consider this example. ■

4.5.2 DENSITY APPROXIMATION APPROACH 1: QUADRATURE INTEGRATION VIA TRANSFORMATION AND CLENSHAW–CURTIS DISCRETE COSINE TRANSFORM QUADRATURE

Here we consider the B-type parameterization, and the starting point will be the evaluation of the integral representation presented in Equation 4.121.

Therefore, to perform density evaluation via the numerical evaluation of the above oscillating integral, one can utilize a weighted quadrature rule, known as the Clenshaw–Curtis quadrature Clenshaw and Curtis (1960) and is presented below. As the Clenshaw–Curtis quadrature only works on finite integral domains, we will first use a generic coordinate transformation which will transform the integral from an infinite interval into a finite one. This is presented according to the following coordinate re-mapping $\theta = C\cot(t)$, where C is a user-specified constant to transform the semi-infinite integral into

$$\begin{aligned} &\frac{1}{2\pi}\int_0^{\infty}\exp\Big\{-\theta^{\alpha}\cos\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big\}\cos\Big[\theta x+\theta^{\alpha}\sin\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big]\,d\theta \\ &= \frac{C}{\pi}\int_0^{\pi}\frac{\sin(t)}{[1-\cos(t)]^2}\exp\Big\{-\Big[C\cot^2\Big(\frac{t}{2}\Big)\Big]^{\alpha}\cos\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big\} \\ &\quad\times\cos\Big[C\cot^2\Big(\frac{t}{2}\Big)x+\Big[C\cot^2\Big(\frac{t}{2}\Big)\Big]^{\alpha}\sin\Big(-\frac{\pi}{2}\beta K(\alpha)\Big)\Big]\,dt. \end{aligned} \tag{4.122}$$

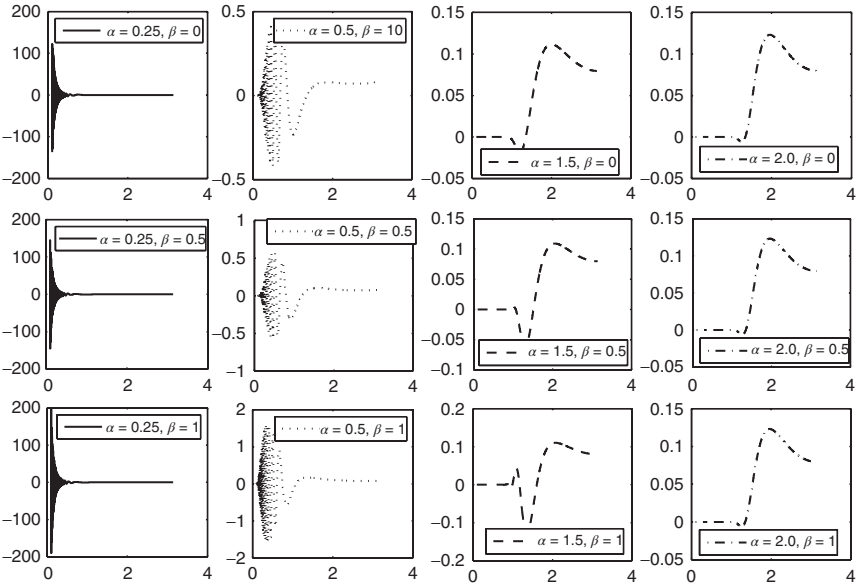

FIGURE 4.6 **Transformed integrand on interval $[0, \pi]$ for values of $\alpha \in \{0.25, 0.5, 1.5, 2\}$ and $\beta \in \{0, 0.5, 1\}$.**

Next one considers the factor multiplying the term $\sin(t)$ in the integrand, making explicit the fact that it may now be considered a function of $\cos(t)$ denoted by $\psi(\cos(t); \alpha, \beta, x)$ and given by

$$\psi(\cos(t); \alpha, \beta, x) = \frac{C}{\pi[1 - \cos(t)]^2} \exp\left\{ -\left[C\cot^2\left(\frac{t}{2}\right)\right]^{\alpha} \cos\left(-\frac{\pi}{2}\beta K(\alpha)\right)\right\}$$
$$\times \cos\left[C\cot^2\left(\frac{t}{2}\right) x + \left[C\cot^2\left(\frac{t}{2}\right)\right]^{\alpha} \sin\left(-\frac{\pi}{2}\beta K(\alpha)\right)\right]. \tag{4.123}$$

We plot this new integrand over the domain of integration in Figure 4.6.

It is evident that under the transformation, the damped oscillatory behaviour is clearly driven primarily by the value of the tail index α. This is less affected by the skewness parameter β. In addition, we note that one can eliminate the singularity at $\theta = 0$ in this integrand, as long as the requirement that the numerator of the integrand plotted goes to zero sufficiently fast as x approaches infinity, and in particular, this function should decay at least as fast as $x - 3/2$. Otherwise if this is not satisfied, one would utilize a different transformation mapping. In addition, we note that from these diagrams, we observe that the integration would be trivially performed without weighting for ranges of $\alpha > 1$.

Furthermore, we note that because this integrand is now explicitly a function of cosines, in which the integrand is given by function $\psi(\cos(t); \alpha, \beta, x)$, this means that we can expand this function in terms of a cosine series given by

$$\psi(\cos(t); \alpha, \beta, x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \frac{2a_{2k}}{1 - (2k)^2} \tag{4.124}$$

with

$$a_k = \frac{2}{\pi}\int_0^{\pi} \psi(\cos(t);\alpha,\beta,x)\cos(kt)dt. \tag{4.125}$$

In order to calculate these cosine coefficients, one is required to again perform an integration for each term. However, the integrals to be evaluated now are of a special form which can exploit the nature of Fourier-series integrations for periodic functions. Therefore, up to the Nyquist frequency $k = N$, these are accurately computed by the $N + 1$ equally spaced and equally weighted points $t_n = \frac{n\pi}{N}$, for integers $n \in \{0, 1, 2, \ldots, N\}$, with the end points weighted by 0.5 to avoid double counting. This would produce the approximation of the cosine series according to the Type-I DCT (see Ahmed *et al.* (1974)) given by

$$a_k \approx \frac{2}{N}\left[\underbrace{\frac{\psi(1;\alpha,\beta,x)}{2} + (-1)^k\frac{\psi(-1;\alpha,\beta,x)}{2}}_{\text{Left and right end points}} + \sum_{n=1}^{N-1}\underbrace{\psi\left(\cos\left[\frac{n\pi}{N}\right];\alpha,\beta,x\right)\cos\left(\frac{nk\pi}{N}\right)}_{\text{Internal points}}\right]$$

for $k \in \{0, 1, 2, \ldots, N\}$.

Hence, with this DCT representation, the solution to the integral is then performed component wise for each term such that one obtains the approximation via the Clenshaw–Curtis quadrature rule (see detailed explanation in Evans and Webster 1999 and Boyd 2001) which in the case presented here for the α-stable density becomes

$$\begin{aligned} f(x;\alpha,\beta;B) &= \frac{1}{2\pi}\mathscr{R}\mathrm{e}\int_0^{\infty}\exp\Big(-\mathrm{i}\theta x - \theta^{\alpha}\exp\Big(-\mathrm{i}\frac{\pi}{2}\beta K(\alpha)\Big)\Big)\,d\theta \\ &\approx \frac{a_0}{2} + \sum_{k=1}^{\infty}\frac{2a_{2k}}{1-(2k)^2}. \end{aligned} \tag{4.126}$$

4.5.3 DENSITY APPROXIMATION APPROACH 2: ADAPTIVE QUADRATURE INTEGRATION VIA FAST FOURIER TRANSFORM (MIDPOINT RULE) AND BERGSTROM SERIES TAIL EXPANSION

We start this section with a brief reminder on how the FFT may be applied to perform numerical inversion of the characteristic function to find the density. In this subsection, we explain how to apply the FFT to evaluate efficiently the discrete Fourier transform (DFT) for the inversion of the characteristic function to find the pdf over a grid of N points. Then we introduce the notion of the fractional FFT for non-integer powers and discuss the properties of this class of transforms for density evaluation.

4.5.3.1 Standard DFT Inversion via the FFT. Consider the generic density given by the integral representation with respect to the characteristic function of random variable X

given by $\Phi_X(\theta)$

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-\mathrm{i}\theta x)\Phi_X(\theta)\, d\theta$$
$$= \frac{1}{\pi}\int_{0}^{\infty} \exp(-\mathrm{i}\theta x)\Phi_X(\theta)\, d\theta. \tag{4.127}$$

First, one discretizes the integral over N equispaced grid points in an interval $[\theta_L, \theta_U]$ using a quadrature rule such as mid-point or trapezoidal to get points $\{\theta_i\}_{i=1}^N$ and the approximation under say the trapezoidal rule

$$\int_a^b h(x)dx = \left(\frac{h(a)+h(b)}{2}\right)\Delta x + \sum_{k=1}^{N-1} h(a+k\Delta x)\Delta x, \tag{4.128}$$

which gives

$$f_X(x) \approx \frac{1}{\pi}\sum_{m=0}^{N-1} \delta_m \exp(-\mathrm{i}\theta_m x)\Phi_X(\theta_m)\Delta\theta, \tag{4.129}$$

where

$$\delta_m = \begin{cases} \frac{1}{2}, & m = 0, \\ 1, & m \neq 0. \end{cases} \tag{4.130}$$

Then set $\nu = \Delta\theta$ and $\theta_m = \nu m$. Next set $x_j = -b + \rho j$ with $\rho = 2\pi/(\nu N)$ as the return grid over which the density will be evaluated in which the parameter b controls the return range. Note that centring the return grid around the origin is easily achieved by selecting $b = \rho N/2$. The corresponding N-point density approximation via the DFT then becomes the series expansion

$$f_X(x_j) \approx \frac{1}{N}\sum_{m=0}^{N-1} f_m \exp\left(-jm\frac{2\pi}{N}\mathrm{i}\right) \tag{4.131}$$

with coefficients given by

$$f_m = \delta_m \frac{N}{\pi}\exp(\mathrm{i}\theta_m b)\Phi_X(\theta_m)\nu. \tag{4.132}$$

This summation has the form required for the FFT to be applied to solve for the density evaluation. In this way, the FFT is an efficient algorithm for computing the discrete Fourier coefficients, via a mapping from $(f_0, f_1, \ldots, f_{N-1})$ to the vector of Fourier coefficients $(d_0, d_1, \ldots, d_{N-1})$ represented by the relationship

$$d_j = \frac{1}{N}\sum_{m=0}^{N-1} f_m \exp\left(-jm\frac{2\pi}{N}\mathrm{i}\right), \quad j = 0, 1, \ldots, N-1, \tag{4.133}$$

when N is an integer power of 2. One would typically utilize the Cooley–Tukey FFT algorithm which recursively decomposes the DFT of any composite size $N - N_1N_2$ into many smaller DFTs.

Remark 4.23 *Note that the use of the FFT to evaluate the DFT coefficients reduces the number of multiplications in the required N summations from an order of 2^{2n} to an order of $n2^{n-1}$.*

4.5.3.2 Fractional Fourier Transforms. Here we consider the fractional Fourier transform (FrFT) which is given in Definition 4.24, see discussions in Ozaktas *et al.* (2001).

Definition 4.24 (Fractional Fourier Transform) *For any real-value constant $\lambda \in \mathbb{R}$, the λ-angle FrFT of a function f (say the characteristic function) is defined by*

$$\mathscr{F}_\alpha[f](u) := \sqrt{1 - \mathrm{i}\cot(\lambda)}\exp\left[\mathrm{i}\pi\left(\mathrm{cosecant}(\lambda)ux - \frac{\cot(\alpha)}{2}x^2\right)\right] f(x)\,dx. \quad (4.134)$$

Remark 4.24 *If one selects $\lambda = \pi/2$ and the function is set to be the characteristic function $f = \Phi_X$, then the FrFT becomes the standard FT that is used to obtain the density.*

There are also analogous discrete fractional Fourier transforms (DFrFT), see discussions in Candan *et al.* (2000), Pei and Yeh (1997) and Pei and Hsue (2009). In these works, the authors were able to consolidate a definition that has the same relation with the DFT as the continuous FrFT has with the ordinary continuous FT. The definition they developed has the following properties:

1. unitarity;
2. index additivity;
3. reduction to the DFT when the order is equal to unity;
4. approximation of the continuous FrFT.

There is, therefore, also a corresponding fractional fast Fourier transform (FrFFT) that can be considered. In the context of the application of the FFT above for solution to the DFT coefficients, if we instead consider the FrFFT, we would be looking at the relationship

$$d_j = \frac{1}{N}\sum_{m=0}^{N-1} f_m \exp(-jm\zeta\mathrm{i}), \quad j = 0, 1, \ldots, N-1, \quad (4.135)$$

for any value of the parameter ζ. By considering the FrFFT approach, we can separate the grid spacings in the domain of the characteristic function arguments θ as characterized by grid spacing parameter ν, from the grid spacings in the domain of the density argument x as characterized by grid spacing parameter ρ, allowing for greater accuracy in the resulting integration in regions of the support of either function where abrupt changes occur. Note that it removes the restriction imposed by the FFT that $\nu\rho = 2\pi/N$.

To understand how the FrFFT applies, consider the vector $\boldsymbol{d} = D(\boldsymbol{f}, \lambda)$ where D is the FrFFT operator on the density coefficients vector $\boldsymbol{f}$. The N-point FrFFT is implemented by using three $2N$-point FFT operations. First, one defines the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ each of which is comprised $2N$ points according to

$$\begin{aligned}\boldsymbol{y} &= [f_0, f_1\exp(\mathrm{i}\pi\lambda), \ldots, f_n\exp(\mathrm{i}\pi n^2\lambda), \ldots, f_{N-1}\exp(\mathrm{i}\pi(N-1)^2\lambda), 0, \ldots, 0],\\ \boldsymbol{z} &= [1, \exp(\mathrm{i}\pi\lambda), \ldots, \exp(\mathrm{i}\pi n^2\lambda), \ldots, \exp(\mathrm{i}\pi(N-1)^2\lambda),\\ &\quad \exp(\mathrm{i}\pi N^2\lambda), \exp(\mathrm{i}\pi(N-1)^2\lambda), \ldots, \exp(\mathrm{i}\pi(n-k)^2\lambda), \ldots, \exp(\mathrm{i}\pi\lambda)], \quad (4.136)\end{aligned}$$

where the last N points of vector $\boldsymbol{y}$ is padded with zeros. Then the FrFFT is defined by the following operation

$$D_k(\boldsymbol{f}, \lambda) = \boldsymbol{w} \odot D_k^{-1}(D_j(\boldsymbol{y}) \odot D_j(\boldsymbol{z})), \tag{4.137}$$

where the $N \times 1$ vector $\boldsymbol{w}$ is given by

$$\boldsymbol{w} = [1, \exp(\mathrm{i}\pi\lambda), \ldots, \exp(\mathrm{i}\pi n^2\lambda), \ldots, \exp(\mathrm{i}\pi(N-1)^2\lambda)], \tag{4.138}$$

and $D_k^{-1}(\cdot)$ is the inverse FFT operator and $\odot$ denotes the element-wise vector multiplication.

The evaluation of the components in the domain $[-a, 0)$ and $[0, a)$ are then performed approximately via the DFT grid approximation via an FFT or an FrFFT as discussed earlier. The other truncation components for the left and right tails on intervals $(-\infty, -a)$ and (a, ∞) can then be approximated as in Luo and Shevchenko (2011), which involves the following details given in Proposition 4.6.

If one considers the integral for the example of the imaginary component such that one first further decomposes the integral,

$$\begin{aligned} f(x; \alpha, \beta, \gamma, \delta; \cdot) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta \\ &= \underbrace{\frac{1}{2\pi} \int_{-\infty}^{\infty} \cos(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta}_{\text{Real part} - \text{denoted } I_1} \\ &\quad - \mathrm{i} \underbrace{\frac{1}{2\pi} \int_{-\infty}^{\infty} \sin(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta}_{\text{Imaginary part} - \text{denoted } I_2}. \end{aligned} \tag{4.139}$$

Now to approximate each integrand, one needs to integrate each of the terms for the real and imaginary components. This will require numerical approximation where we first consider separation of the integral to a bounded domain given by

$$\begin{aligned} I_1 &= \frac{1}{2\pi} \int_{-\infty}^{-a} \cos(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta \\ &\quad + \frac{1}{2\pi} \int_{-a}^{0} \cos(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta \\ &\quad + \frac{1}{2\pi} \int_{0}^{a} \cos(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta \\ &\quad + \frac{1}{2\pi} \int_{a}^{\infty} \cos(\theta x) \Phi_X(\theta; \alpha, \beta, \gamma, \delta; \cdot) d\theta. \end{aligned} \tag{4.140}$$

Now we consider the remainder integrals over intervals $(-\infty, -a)$ and (a, ∞) that were not obtained from the DFT because of truncation. Then we can state the following results in Proposition 4.6 for approximation of the remainder.

Proposition 4.6 *If the integrand $\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)$ satisfies the following two conditions:*

C1. The following integral exists

$$\frac{1}{2\pi}\int_a^{\infty}\sin(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta,$$

C2. The following derivatives exist

$$\frac{d^k}{d\theta^k}\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)$$

and converge to $\to 0$ as $k\to\infty$ for all $k\in\mathbb{Z}^+$,

then a simple approximation to the truncation error is given by selecting $a = N\pi$ for some large integer N to produce the approximation given by

$$\frac{1}{2\pi}\int_{N\pi}^{\infty}\sin(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \approx (-1)^N\Phi_X(N\pi;\alpha,\beta,\gamma,\delta;\cdot). \tag{4.141}$$

Furthermore, as derived in Luo and Shevchenko (2011), the accuracy of this approximation can be obtained assuming further the following condition:

C3. the derivatives $\frac{d^{2k}}{d\theta^{2k}}\Phi_X(a=\frac{N}{\pi};\alpha,\beta,\gamma,\delta;\cdot)=0$ for all $k\in\mathbb{Z}^+$,

which gives the truncation approximation error, denoted by e_T, according to

$$\begin{aligned} e_T = &\sum_{i=1}^{k-1}(-1)^{N+i}\frac{d^{2i}}{d\theta^{2i}}\Phi_X(N\pi;\alpha,\beta,\gamma,\delta;\cdot) \\ &+(-1)^k\int_{N\pi}^{\infty}\frac{d^{2k}}{d\theta^{2k}}\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)\sin(\theta)d\theta, \end{aligned} \tag{4.142}$$

which has the asymptotic property as $N\to\infty$ according to

$$\frac{1}{2\pi}\int_{N\pi}^{\infty}\sin(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \sim \sum_{i=0}^{\infty}(-1)^{N+i}\frac{d^{2i}}{d\theta^{2i}}\Phi_X(N\pi;\alpha,\beta,\gamma,\delta;\cdot). \tag{4.143}$$

The same holds for the negative tail approximation.

4.5.3.3 Adaptive Quadrature Integration via Fast Fourier Transform (Midpoint Rule) and Bergstrom Series Tail Expansion. One can now apply either the FFT or the FrFFT solutions to the DFT rule combined with a quadrature rule such as the midpoint rule considered below. This second numerical solution proceeds as follows, where one considers $X\sim S(\alpha,\beta,\gamma,\delta,\cdot)$ and performs the evaluation of the density and distribution function pointwise according to the solution to the inversion of the FT integral equations, where if the characteristic function parameterization of interest is integrable, then the distribution $F(x;\alpha,\beta,\gamma,\delta;\cdot)$ is absolutely continuous and the severity distribution would have a

probability distribution $f(x;\alpha,\beta,\gamma,\delta;\cdot)$ found as the solution to the integral w.r.t. the characteristic function $\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)$

$$\begin{aligned} f(x;\alpha,\beta,\gamma,\delta;\cdot) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-\mathrm{i}\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \\ &= \underbrace{\frac{1}{2\pi}\int_{-\infty}^{\infty}\cos(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta}_{\text{Real part}-\text{denoted by } I_1} \\ &\quad - \mathrm{i}\underbrace{\frac{1}{2\pi}\int_{-\infty}^{\infty}\sin(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta}_{\text{Imaginary part}-\text{denoted by } I_2}. \end{aligned} \tag{4.144}$$

Now to approximate each integrand, one needs to integrate each of the terms for the real and imaginary components. This will require numerical approximation where we first consider separation of the integral to a bounded domain given by

$$\begin{aligned} I_1 &= \frac{1}{2\pi}\int_{-\infty}^{-a}\cos(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \\ &\quad + \frac{1}{2\pi}\int_{-a}^{0}\cos(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \\ &\quad + \frac{1}{2\pi}\int_{0}^{a}\cos(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta \\ &\quad + \frac{1}{2\pi}\int_{a}^{\infty}\cos(\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta. \end{aligned} \tag{4.145}$$

The evaluation of the components in the domain $[-a,0)$ and $[0,a)$ are then performed approximately via the DFT grid approximation via an FFT as discussed in Cruz *et al.* (2015) and detailed in (Doganoglu and Mittnik 1998, section 2). Note that the FFT is an efficient approach to the calculation of the DFT coefficients when the number of discretization points is a power of 2. The integrals over the domain $(-\infty,-a)$ and (a,∞) are then approximated by series representations as detailed below.

The steps required are detailed in Example 4.7.

EXAMPLE 4.7 FFT for DFT Coefficients

Consider the standardized stable random variables characteristic function $\Phi_X(\theta;\alpha,\beta,\gamma=1,\delta=0;\cdot)$ which is to be approximated over a finite interval $[-a,a]$ according to

$$f(x;\alpha,\beta;\cdot) \approx \frac{1}{2\pi}\int_{-a}^{a}\exp(-\mathrm{i}\theta x)\Phi_X(\theta;\alpha,\beta,\gamma,\delta;\cdot)d\theta. \tag{4.146}$$

Then one may use any number of basic quadrature approximations such as rectangular rules or Simpson rule given by say the midpoint rule or the Simpson rule:

- **Midpoint rule:** $\int_a^b f(x)dx \approx f(x_m)(b-a)$ with $x_m = (b-a)/2$.
- **Simpson rule:** $\int_a^b f(x)dx \approx \frac{1}{6}(f(a)+4f(x_m)+f(b))(b-a)$ with $x_m = (b-a)/2$.

So if one takes, for example, the midpoint rule, the approximation of the integral over domain $[-a, a]$ is given by FFT as follows. Consider $N = 2^m \in \mathbb{N}$ grid points which are equispaced, with spacing $\Delta = 2a/N$, and given by $\{\theta_k\}_{k=0}^N$ and midpoints of these intervals by $\{\theta_k^{(m)}\}_{k=0}^N$

$$\theta_k = -a + k\Delta, \quad k = 0, \ldots, N,$$
$$\theta_k^{(m)} = 0.5(\theta_k + \theta_{k+1}), \quad k = 0, \ldots, N.$$

Furthermore, define the function for the integrand evaluated at these points by the discrete sequence $\{y_j\}_{j=0}^{N-1}$ given by

$$y_j = (-1)^j \Phi_X(\theta_k^{(m)}; \alpha, \beta, \gamma, \delta; \cdot), \quad k = 0, \ldots, N-1.$$

Now the DFT will map the sequence $\{y_j\}_{j=0}^{N-1} \in \mathbb{C}^N$ to a new sequence $\{Y_j\}_{j=0}^{N-1} \in \mathbb{C}^N$ where

$$Y_j = \sum_{j=0}^{N-1} y_j \exp\left(-\mathrm{i}l\frac{2\pi j}{N}\right) = DFT\{y_j\}_{j=0}^{N-1}.$$

To evaluate this DFT efficiently, we may use the FFT as follows. Define the coefficients $\{C_k\}_{j=0}^{N-1} \in \mathbb{C}^N$ by the following

$$C_k = \frac{2a}{N}(-1)^k \exp\left(-\mathrm{i}\frac{\pi}{N}k\right), \quad k = 0, \ldots, N,$$

then for the points $\{x_k\}_{k=0}^{N-1}$ given by

$$x_k = -\frac{N\pi}{2a} + \frac{\pi}{a}k, \quad k = 0, \ldots, N$$

one has the following application of the 'midpoint rule' under the DFT given by

$$\sum_{j=0}^{N-1} \Phi_X(\theta_k^{(m)}; \alpha, \beta, \gamma, \delta; \cdot) \exp(-\mathrm{i}\theta_j^{(m)} x_k)(\theta_{j+1} - \theta_j) = C_k DFT\{y_j\}_{j=0}^{N-1},$$

which can then be efficiently solved via an FFT. ■

Remark 4.25 *As noted in Menn and Rachev (2006), the selection of the evaluation range, as specified by a, is critical to the accuracy of the method. However, such a range will differ depending on the parameters of the stable distribution, especially the tail index α. To address this, one could utilize the calibrated FFT-based density approximation proposed in Menn and Rachev (2006, section 4).*

The other truncation components for the left and right tails represent the error in the approximation because of truncation; these components can be approximated using a series expansion attributed to Bergstrom (1953) and utilized in the context of approximation of the

tails in the FFT evaluation in the papers of Menn and Rachev (2006). This series expansion for the standardized α-stable severity model is given for the $S1$ parameterization of Nolan in Lemma 4.6 and also given in a similar form in Zolotarev (1986, corollary 2, p. 94) for the B-type parameterization. This series expansion can then be utilized to some order n to approximate the remainder tails for the density.

Lemma 4.6 (Bergstrom Series Expansion) *Consider the loss random variable $X \sim S_\alpha(\beta, 0, 1; S1)$, then, in the parameter range $1 < \alpha \leq 2$, the density for the standardized stable random variable $f(x; \alpha, \beta; S1)$ has the asymptotic expansion given as $x \to \infty$ by*

$$f(x; \alpha, \beta; S1) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} c a_k (cx)^{-k\alpha - 1} \tag{4.147}$$

with

$$\begin{aligned}
\beta^* &= \frac{2}{\pi(\alpha - 2)} \arctan\left(-\beta \tan \frac{\pi\alpha}{2}\right), \\
c &= \cos\left(\frac{\pi\beta^*(\alpha - 2)}{2}\right)^{1/\alpha}, \\
a_k &= (-1)^{k-1} \frac{\Gamma(1 + k\alpha)}{k!} \sin\left(\frac{k\pi}{2}(\alpha + \beta^*(\alpha - 2))\right).
\end{aligned}$$

In addition, a similar asymptotic series expansion for the distribution function of the B-type standardized model was obtained in Zolotarev (1986, p. 95), see Proposition 4.7.

Proposition 4.7 (Zolotarev's Asymptotic Series Expansions Type B) *Consider the standardized B-type α-stable random variable, standardized such that $\gamma = 1$ and $\delta = 0$. The density can be evaluated asymptotically with $x \to \infty$ by*

$$f_X(x; \alpha, \beta; B) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{k!} P_k(\ln x) x^{-k-1} \tag{4.148}$$

for $\alpha = 1$ and $\beta \in (-1, 1]$ and with

$$\begin{aligned}
P_k(x) &= \sum_{l=0^k} r_{lk} x^l, \\
r_{lk} &= \sum_{m=l}^{k} C_m^k \, C_l^m (-1)^{m-l} \Gamma^{(m-l)}(1 + n) \beta^m \left(\frac{\pi}{2}(1 + \beta)\right)^{k-m} \sin\left(\frac{\pi}{2}(k - m)\right),
\end{aligned}$$

where $\Gamma^{(m)}(x)$ is the m-th derivative of the Gamma function.

4.5.4 DENSITY APPROXIMATION APPROACH 3: TRUNCATED POLYNOMIAL SERIES EXPANSIONS

A third alternative which utilizes the FFT solutions above in a more efficient manner when evaluation of the pdf must be performed over a large set of points is based on polynomial expansions provided in Doganoglu and Mittnik (1998, section 3). The approach they develop

is simple and involves considering a finite series approximation to the density of the α-stable severity. To achieve the series representation, they take the standardized loss random variable $Z \sim S(\alpha, \beta, 1, 0; S1)$ under Nolan's $S1$ parameterization and they first transform the argument of the stable loss by $X = Z - \beta \tan(\pi\alpha/2)$, such a transformation shifts the mass of the distribution to ensure that for almost all values of α and β, the probability mass lies in an interval around the origin. With this transformed random variable, they consider the polynomial series representation over a grid of N values for Z given by $\{z_i\}_{i=1}^N$, then for each z_i, one obtains a value of $x_i = z_i - \beta \tan(\pi\alpha/2)$ to get the set of grid points $\{x_i\}_{i=1}^N$.

They also consider an array of $\{\alpha_k\}_{k=1}^K$ values for α and an array of $\{\beta_l\}_{l=1}^L$ values for β to produce the series representation given by

$$f_X(x; \alpha, \beta, 1, 0; S1) \approx \sum_{k=k_1}^{k_K} \sum_{l=l_1}^{l_L} \sum_{m=m_1}^{m_M} c_{klm} \alpha_k \beta_l x^m, \tag{4.149}$$

where one finds the coefficients c_{klm} by solving a linear system of equations for combinations of (x, α, β). Then for each (α_i, β_j) combination, one has a grid of N values of $x_{1:N}$ which will have density values in a vector $\boldsymbol{f}_{ij} = [f_{1ij}, \ldots, f_{Nij}]$ that are calculated via an FFT procedure, as discussed previously in Approach 2.

Then for each power of x, define the matrix D_{ij} for a pair of parameters (α_i, β_j) which has N rows corresponding to the grid points of the support of random variable X's distribution and M columns for the powers of x. Note that each entry in the $N \times M$ matrix of powers of x values is also a function of the α_i and β_j values because of the transform (hence the ij subscript), so there will actually be a matrix $N \times M$ for each pair of parameter values (α_i, β_j). This means that the evaluation of the density vector $\boldsymbol{f}_{ij}$ is then given by

$$\boldsymbol{f}_{ij} := f(x_{1:N}; \alpha_i, \beta_j; S1) \approx D_{ij}[\mathbb{I}_M \otimes (\boldsymbol{a}_i \otimes \boldsymbol{b}_j)]\boldsymbol{c}, \tag{4.150}$$

with $\mathbb{I}_M$ the $M \times M$ identity matrix, $\boldsymbol{a}_i$ the $1 \times K$ vector of powers of α_i, $\boldsymbol{b}_j$ the $1 \times L$ vector of powers of β_j and $\boldsymbol{c}$ the $KLM \times 1$ vector of coefficients c_{klm}. Hence, one can then solve this system of equations for the coefficients $\boldsymbol{c}$ using say least squares methods. However, the authors note that there may be numerical instability if one attempts a naive solution of this form because of potential for the $NT \times KLM$ matrix $D_{ij}[\mathbb{I}_M \otimes (\boldsymbol{a}_i \otimes \boldsymbol{b}_j)]$ to be ill-conditioned. Hence, they recommend a two-stage numerical solution which involved the following two steps:

Stage 1. For each pair of parameter values (α_i, β_j), fit an Mth order polynomial in x using the FFT values for the density by solving

$$\boldsymbol{f}_{ij} \approx D_{ij} \boldsymbol{c}_{ij}^{(1)}, \tag{4.151}$$

where $\boldsymbol{c}_{ij}^{(1)}$ denotes the $M \times 1$ vector of coefficients at stage (1).

Stage 2. Fit a polynomial in α and β powers by polynomial interpolation for the surface values of $\{\boldsymbol{c}_{ij}^{(1)}\}$ which is an $M \times 1$ vector to obtain the second-stage coefficients which are the solution to

$$c_{ijm}^{(1)} \approx \sum_{k=k_1}^{k_K} \sum_{l=l_1}^{l_L} c_{klm}^{(2)} \alpha_k \beta_l \tag{4.152}$$

for $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, L\}$.

This two-stage procedure for the estimation of the coefficients helps to avoid ill-conditioning of the matrix in the system of equations that would occur if a Newton-type solution technique were applied.

4.5.5 DENSITY APPROXIMATION APPROACH 4: REPARAMETERIZATION

A fourth alternative, developed to overcome the challenge associated with numerical integration when there are oscillations in the integrands, as present in the solutions discussed above, is to consider the numerical integration of the stable characteristic function that was provided in Zolotarev (1986, theorem 2.2.3). It was recognized that one could utilize an alternative parameterization which has the feature that the integrands do not oscillate infinitely many times. This would involve work with the B-type parameterization transformed according to the following representations provided by Zolotarev, given in Theorem 4.18.

Theorem 4.18 *The density for the standard stable distribution with $\gamma = 1$ and $\delta = 0$ under the B-type parameterization can be evaluated using the following convergent series representation*

$$f_X(x;\alpha,\beta;B) = \begin{cases} \dfrac{\alpha|x|^{\frac{1}{\alpha-1}}}{2|1-\alpha|}\displaystyle\int_{-\theta^*}^{1} U_\alpha(\varphi,\theta^*)\exp\left(-|x|^{\alpha/(\alpha-1)}U_\alpha(\varphi,\theta^*)\right)d\varphi, \\ \qquad\qquad \text{if } \alpha \neq 1, x \neq 0, |\beta| \leq 1, \\ \dfrac{1}{2|\beta|}\exp\left(-\dfrac{x}{\beta}\right)\displaystyle\int_{-1}^{1} U_1(\varphi,\beta)\exp\left(-\exp\left(-\dfrac{x}{\beta}\right)U_1(\varphi,\beta)\right)d\varphi, \\ \qquad\qquad \text{if } \alpha = 1 \text{ and } \beta \neq 0, \end{cases} \tag{4.153}$$

with the following functions given by

$$\theta^* = \theta \text{sgn}(x), \quad \theta = \beta\frac{K(\alpha)}{\alpha},$$

$$U_\alpha(\varphi,\theta) = \left(\frac{\sin\left(\frac{\pi}{2}\alpha(\varphi+\theta)\right)}{\cos\left(\frac{\pi}{2}\varphi\right)}\right)^{\frac{\alpha}{1-\alpha}} \frac{\cos\left(\frac{\pi}{2}\left((\alpha-1)\varphi+\alpha\theta\right)\right)}{\cos(\frac{\pi}{2}\varphi)},$$

$$U_1(\varphi,\beta) = \frac{\pi}{2}\frac{(1+\beta\varphi)}{\cos\left(\frac{\pi}{2}\varphi\right)}\exp\left(\frac{\pi}{2}\left(\varphi+\frac{1}{\beta}\right)\tan\left(\frac{\pi}{2}\varphi\right)\right).$$

Using this transformed integral representation, one can then apply any form of quadrature method to solve the integral to a desired accuracy. For instance, we would recommend the Gaussian–Hermite quadrature approach or the Gauss–Kronrod quadrature approach, which has a similar accuracy to the DCT quadrature presented in Approach 1, see detailed discussions on how to implement such a quadrature rule in Laurie (1997) and Calvetti *et al.* (2000). Note that these integrands are then free from oscillations as presented in Figure 4.7.

Of course, one can also find alternative ways to present the density for the stable loss random variable such as series expansions which can be truncated. Before presenting the series representation results that are known for the density, it is useful to recall the following results that are known regarding the bounds on the absolute values of the derivatives of the α-stable

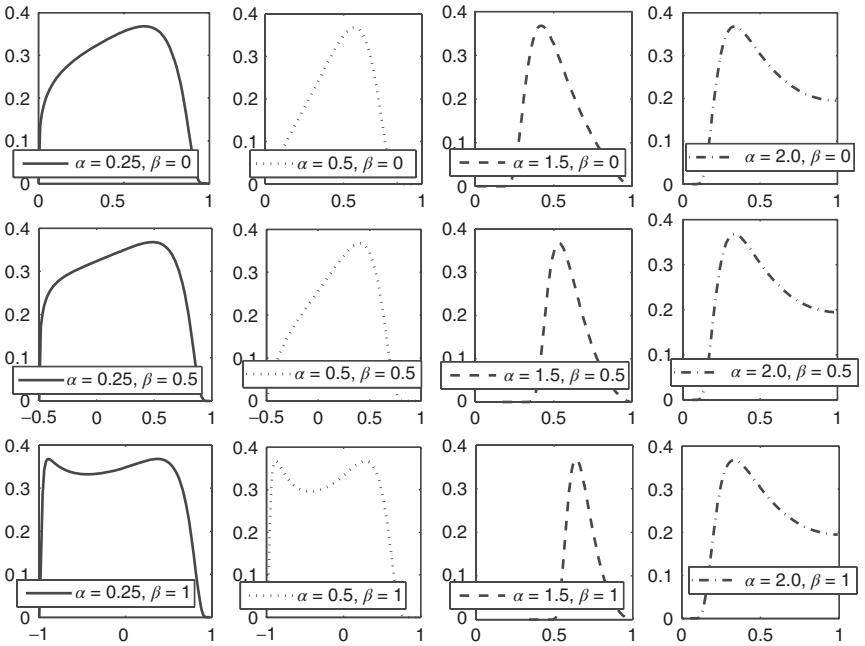

FIGURE 4.7 **Transformed integrands to remove oscillatory behaviour for values of** $\alpha \in \{0.25, 0.5, 1.5, 2\}$ **and** $\beta \in \{0, 0.5, 1\}$.

density under the A-type parameterization in Proposition 4.8, due to Zolotarev (1986, p. 13). Note that these derivative bounds are useful when considering the behaviour and convergence properties of series representations as well as later when performing gradient-based methods for the parameter estimation.

Proposition 4.8 *The modulus of the A-type parameterizations characteristic function is trivially obtained as*

$$|\mathbb{E}[\exp(\mathrm{i}\theta X)]| = \begin{cases} \exp\left(-\gamma|\theta|^{\alpha}\right), & \alpha \neq 1, \\ \exp(-\gamma|\theta|), & \alpha = 1, \end{cases} \tag{4.154}$$

from which it then follows that the corresponding distribution and density function $S_{\alpha}(x; \beta, \gamma, \delta; A)$ *exist and derivatives can be uniformly bounded over their support according to (see Property I.20 in Zolotarev 1986)*

$$\begin{aligned} &\frac{d^n}{dx^n}|S_{\alpha}(x; \beta, \gamma, \delta; A)| \\ &= \begin{cases} \dfrac{d^n}{dx^n}\left|\dfrac{1}{2\pi}\displaystyle\int \exp\left(\mathrm{i}\delta_A\gamma\theta - \gamma|\theta|^{\alpha}\left\{1 - \mathrm{i}\beta\mathrm{sign}(\theta)\tan\left(\frac{\pi\alpha}{2}\right)\right\}\right)d\theta\right|, & \alpha \neq 1, \\ \dfrac{d^n}{dx^n}\left|\dfrac{1}{2\pi}\displaystyle\int \exp\left(\mathrm{i}\delta_A\gamma\theta - \gamma|\theta|\left\{1 + \mathrm{i}\beta\mathrm{sign}(\theta)\frac{2}{\pi}\ln|\theta|\right\}\right)d\theta\right|, & \alpha = 1, \end{cases} \end{aligned}$$

$$\leq \begin{cases} \frac{1}{2\pi}\int |\theta|^n \exp\left(-\gamma|\theta|^\alpha\right) d\theta, & \alpha \neq 1, \\ \frac{1}{2\pi}\int |\theta|^n \exp(-\gamma|\theta|) d\theta, & \alpha = 1, \end{cases}$$

$$= \begin{cases} \frac{\Gamma\left((n+1)/\alpha\right)}{\pi\alpha}\gamma^{-(n+1)/\alpha}, & \alpha \neq 1, \\ \frac{\Gamma(n+1)}{\pi}\gamma^{-(n+1)}, & \alpha = 1. \end{cases} \tag{4.155}$$

4.5.6 DENSITY APPROXIMATION APPROACH 5: INFINITE SERIES EXPANSION DENSITY AND DISTRIBUTION REPRESENTATIONS

The fifth alternative is to consider representing the distribution according to a series expansion. Examples of such expansions have already been presented for asymptotic tail expansions; in the following section, we extend this discussion to the whole support. To achieve this, there are numerous results in this regard, for example, the series expansion representation given in Lemma 4.7.

Lemma 4.7 (α-Stable Severity Density and Distribution Representations) *Consider the standardized B-Type α-stable random variable, standardized such that $\gamma = 1$ and $\delta = 0$. Then the density function can be evaluated pointwise according to the following series expansions (Zolotarev 1986, Equation 2.4.6, p. 89):*

$$f_X(x;\alpha,\beta,1,0;B)$$

$$= \begin{cases} \frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(1+n/\alpha)}{\Gamma(n+1)}\sin\left(n\pi(1+\beta K(\alpha))/(2\alpha)\right)x^{n-1}, \\ \qquad\qquad \text{if } \alpha > 1, \beta \in [-1,1], x \in \mathbb{R}, \\ \frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1} n b_n x^{n-1}, \\ \qquad\qquad \text{if } \alpha = 1, \beta \in (0,1], x \in \mathbb{R}, \\ \frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(n\alpha+1)}{\Gamma(n+1)}\sin\left(n\pi\alpha(1+\beta K(\alpha))/(2\alpha)\right)x^{-n\alpha-1}, \\ \qquad\qquad \text{if } \alpha < 1, \beta \in [-1,1], x \in \mathbb{R}^+, \end{cases} \tag{4.156}$$

where the coefficients b_n are given by

$$b_n = \frac{1}{\Gamma(n+1)}\int_0^\infty \exp(-\beta u \ln u) u^{n-1} \sin\left[(1+\beta)u\frac{\pi}{2}\right] du. \tag{4.157}$$

In Bergstrom (1953), where it was shown that all continuous stable distributions can be written in terms of infinite series expansions in terms of elementary functions, there have been specific examples created. There are the series expansion representations developed in Metzler and Klafter (2000); they transform the parameters and provide the stable densities in terms of the Fox H-functions, see Fox (1961). Other reparameterizations to obtain representations include Hoffmann-Jørgensen (1993), who obtain representations with respect to incomplete hypergeometric functions. However, all these results involve making particular choices for the index of stability parameter α. In the following, we also present a specific generalization of these results developed in Zolotarev (1986) which considers only the restriction that $\beta = 1$,

the perfectly skewed families of α-stable models, ideal for OpRisk modeling. Hence, we first present the general sub-family representations for perfectly skewed examples before presenting some symmetric representations for the severity models, which will need to be truncated for positive support.

In the special case in which one considers the α-stable severity distribution which is perfectly skewed distribution with strictly positive support, as of relevance to the results presented in Chapter 6, the following representations due to Zolotarev (1986, theorem 2.4.4) are available. These representations hold for severity distribution models for the standardized stable models in which $\alpha \in (0, 2)$, $\beta = 1$, $\gamma = 1$ and $\delta = 0$ and are formed in terms of an infinite basis function series expansion, based on generalized Laguerre polynomials, see Definition 4.25.

Definition 4.25 (Laguerre Polynomials) *The generalized Laguerre polynomials are defined as the solutions to the second-order linear differential equation, for integers n according to*

$$xy'' + (\alpha + 1 - x)y' + ny = 0.$$

The solutions form a sequence of polynomials $L_0, L_1, \ldots$ which are also an orthonormal sequence given by several different representations. First, it will be useful to consider the case of the standard Laguerre polynomials in which $\alpha = 0$ and then the following recursive relationship for their definition is considered for $x \in \mathbb{R}$ and for all $k \in \mathbb{N}^+$

$$\begin{aligned} L_0(x) &= 1, \\ L_1(x) &= 1 - x, \\ L_{k+1}(x) &= \frac{1}{k+1}((2k + 1 - x)L_k(x) - kL_{k-1}(x)). \end{aligned} \tag{4.158}$$

Now the generalized Laguerre polynomials, for cases in which $\alpha \neq 0$, are given by

$$\begin{aligned} L_n^{(\alpha)}(x) &= \frac{x^{-\alpha}\exp(x)}{n!}\frac{d^n}{dx^n}(\exp(-x)x^{n+\alpha}) \\ &= \sum_{i=0}^{n}(-1)^i \binom{n+\alpha}{n-i}\frac{x^i}{i!}, \end{aligned}$$

or for $x \in \mathbb{R}$ according to the special functions known as the confluent hypergeometric functions or Kummer's functions by

$$\begin{aligned} L_n^{(\alpha)}(x) &= \binom{n+\alpha}{n} M(-n, \alpha + 1, x) \\ &= \frac{(-1)^n}{n!}U(-n, \alpha + 1, x), \end{aligned} \tag{4.159}$$

where one defines

$$M(a, b, z) = \sum_{n=0}^{\infty}\frac{(a)_{(n)}z^n}{(b)_{(n)}n!},$$

$$U(a, b, z) = \frac{\Gamma(1-b)}{\Gamma(a-b+1)}M(a, b, z) + \frac{\Gamma(b-1)}{\Gamma(a)}z^{1-b}M(a-b+1, 2-b, z),$$

with Pochhammer symbols $(a)_{(n)} = a(a + 1)(a + 2)\cdots(a + n - 1)$.

These generalized Laguerre polynomials can then be utilized to obtain a series expansion for the α-stable severity density as follows.

Lemma 4.8 (α-Stable Strictly Positive Support Severity Density Representations) *The standardized perfectly skewed α-stable severity density $S_\alpha(x; 1, 1, 0; B)$ can be represented in the heavy-tailed cases by the following Laguerre polynomial series expansion for $x > 0$, see Zolotarev (1986, theorem 2.4.4)*

$$f_X(x;\alpha,1,1,0;B) = \begin{cases} x^{(1-\alpha^2)/\alpha} \exp(-x^{-\alpha}) \sum_{n=0}^{\infty} k_n^{(s)}(\alpha)\, L_n^{(s)}(x^{-\alpha}), & \forall x > 0, 0 < \alpha < 1, \\ \frac{1}{x} \exp(-x) \sum_{n=0}^{\infty} k_n^{(s)}(1/\alpha)\, L_n^{(s)}(x), & \forall x > 0, 1 < \alpha < 2, \end{cases} \tag{4.160}$$

with the coefficient functions defined by

$$k_n^{(s)}(\alpha) = \alpha \left(\frac{\Gamma(n+1)}{\Gamma(n+1+s)} \right)^{\frac{1}{2}} \sum_{m=0}^{n} \frac{(-1)^m \Gamma(1+s+n)}{\Gamma(m+1)\Gamma(n-m+1)\Gamma(1+\alpha(s+m))}.$$

As an example of one of the special cases of this representation, in which the series representation takes an even simpler closed-form expression in terms of Whittaker functions, see Example 4.8.

EXAMPLE 4.8

If one considers the α-stable severity distribution model with perfect support, then the choice of $\alpha = 2/3$ in the standardized perfectly skewed α-stable severity density $S_\alpha(x; 1, 1, 0; B)$ presented in Lemma 4.8 above reduces to the following expression

$$f_X(x; 2/3, 1, 1, 0; B) = \frac{x^{-1}}{\sqrt{3\pi}} \exp\left(-\frac{2}{27} x^{-2} \right) W_{1/2,1/6}\left(\frac{4}{27} x^{-2} \right), \tag{4.161}$$

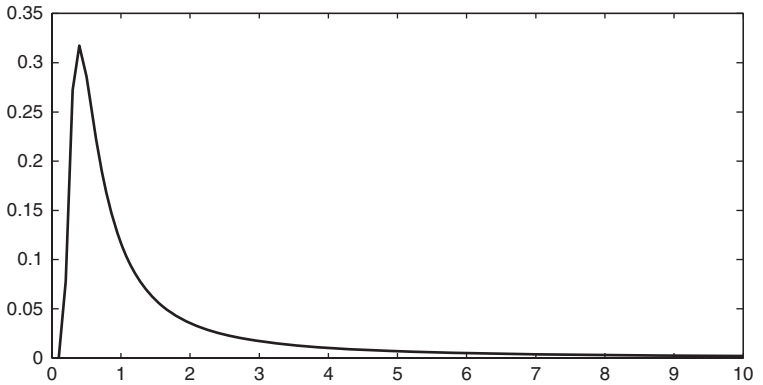

FIGURE 4.8 **Strictly positive support B-type stable density function based on the special function given by the Whittacker function representation with $\alpha = 2/3$, $\beta = 1$, $\gamma = 1$ and $\delta = 0$.**

where the Whittaker function (see Abramowitz and Stegun 1972, p. 505) is given in terms of the confluent hypergeometric functions according to

$$W_{k,m}(x) = \exp\left(-\frac{x}{2}\right) x^{(m+1)/2} U\left(\frac{1}{2} + m - k, 1 + 2m; x\right). \tag{4.162}$$

In Figure 4.8, we present an example of the density expansion above for the case corresponding to the Whittaker function example. ■

In addition, one can also show two additional special cases of such expansions for the symmetric case, which can be truncated to positive supports for application in OpRisk. First, we define the relevant special functions given by hypergeometric functions in Definition 4.26.

Definition 4.26 *Consider the function on the domain $|x| < 1$, then the hypergeometric function is given by the convergent polynomial series representation*

$${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x) = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{x^n}{n!} \tag{4.163}$$

with the Pochhammer symbols given by

$$(p)_n = \begin{cases} 1, & \text{if } n = 0, \\ p(p+1)\cdots(p+n-1), & \text{if } n > 0. \end{cases} \tag{4.164}$$

Using these special functions, we obtain the specific representation given next.

Lemma 4.9

1. Finite Mean Severity Model ($\alpha = 3/2$). *The Holtsmark density representation of Garoni and Frankel (2002) is given for standardized symmetric stable loss random variable according to the analytic density represented according to hypergeometric distributions as follows:*

$$\begin{aligned} f_X(x) = {} & \frac{1}{\pi}\Gamma(5/3)\,{}_2F_3\left(\frac{5}{12}, \frac{11}{12}; \frac{1}{3}, \frac{1}{2}, \frac{5}{6}; -\frac{2^2x^6}{3^6}\right)\frac{x^2}{3\pi^2} \\ & \times {}_3F_4\left(\frac{3}{4}, 1, \frac{5}{4}; \frac{2}{3}, \frac{5}{6}, \frac{7}{6}, \frac{4}{3}; -\frac{2^2x^6}{3^6}\right) \\ & + \frac{7x^4}{3^4\pi}\Gamma(4/3)\,{}_2F_3\left(\frac{13}{12}, \frac{19}{12}; \frac{7}{6}, \frac{3}{2}, \frac{5}{3}; -\frac{2^2x^6}{3^6}\right), -\infty < x < \infty, \end{aligned} \tag{4.165}$$

with $\alpha = \frac{3}{2}, \beta = 0, \gamma = 1, \delta = 0$.

2. Infinite Mean Severity Model ($\alpha = 2/3$). *The Whittaker function density representation Uchaikin and Zolotarev (1999) is given for a random variable denoted by $\widetilde{X}_{2/3}$ which has an analytic density which is represented according to the Whittaker functions, see Abramowitz and Stegun (1965), as follows:*

$$f_X(x) = \frac{1}{2\sqrt{3\pi}}|x|^{-1}\exp\left(\frac{2}{27}x^{-2}\right) W_{-1/2;1/6}\left(\frac{4}{27}x^{-2}\right), -\infty < x < \infty, \tag{4.166}$$

with $\alpha = \frac{2}{3}, \beta = 0, \gamma = 1, \delta = 0$ *and*

$$W_{\lambda;\mu}(z) = \frac{z^{\lambda}\exp(-z/2)}{\Gamma(\mu-\lambda+1/2)}\int_0^{\infty}\exp(-t)t^{\mu-\lambda-1/2}\left(1+\frac{t}{z}\right)^{\mu-\lambda-1/2}dt.$$

Other authors have also derived similar series expansion results such as those of Hall (1984) and Holt and Crow (1973) where it is shown that when the tail index is in the range $\alpha \in (1,2)$, the density function of an α-stable distribution can be written as specified in Lemma 4.10.

Lemma 4.10 *Given a univariate stable random variable* $X \sim S(x;\alpha,\beta,\gamma,\delta;0)$ *with tail index in the range* $\alpha \in (1,2)$*, the following series expansion of the density function applies:*

$$f(x;\alpha,\beta,1,0) = \frac{1}{\pi}\sum_{i=1}^{\infty}\frac{(-x)^{i-1}\Gamma(1+i/\alpha)}{i!r^i}\sin\left(\frac{i\pi}{2\alpha}(\alpha+s)\right)\ x>0, \tag{4.167}$$

where $r = \left(1+\tan^2(\pi s/2)\right)^{-1/(2\alpha)}$ *and* $s = -\frac{2}{\pi}\arctan(\beta\tan(\pi\alpha/2))$.

EXAMPLE 4.9

In this example, the aim is to study how the stable density approximation given by the series representation

$$f(x;\alpha,\beta,1,0) = \frac{1}{\pi}\sum_{i=1}^{\infty}\frac{(-x)^{i-1}\Gamma(1+i/\alpha)}{i!r^i}\sin\left(\frac{i\pi}{2\alpha}(\alpha+s)\right)\ x>0$$

behaves for a range of parameter values for $\alpha \in \{1.9, 1.6, 1.4\}$ and $\beta = \{0, 0.5, 0.9\}$. To undertake this study, we wish to demonstrate the behaviour of say the first 20 summands over a grid of values of $x \in \{0.1, 0.5, 1, 3\}$. The intention is to illustrate that as one moves into the tails of the distribution, utilization of this series representation to accurately evaluate the density requires an increasing number of terms for an accurate evaluation. In general, the number of terms will also be highly dependent on the parameter values; as α decreases and the tails get heavier at any given point, one requires, in general, a greater number of terms for accurate evaluation. We first present the case of $\beta = 0$ with a range of α values for four grid points for x in Figure 4.9.

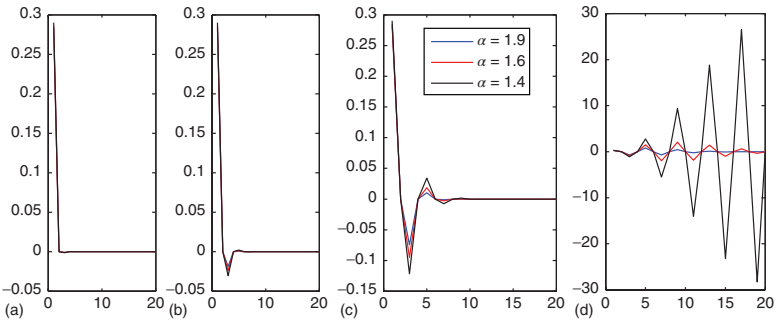

FIGURE 4.9 **Study of first 20 summand terms of the stable density series expansion for a range of parameter values at three different x locations. (a) $x = .1$, (b) $x = 0.5$, (c) $x = 1$ and (d) $x = 3$.** ***See insert for color representation of this figure.***

These results demonstrate that as one evaluates the tails of the severity distribution using this series expansion; if the tail index is decreasing, you must increase the number of summands. In general, the guide would be to keep increasing terms n in the summation at a given point x, until a tolerance is satisfied such as

$$|\widehat{f}_n(x;\alpha,\beta) - \widehat{f}_{n-1}(x;\alpha,\beta)| < \varepsilon, \tag{4.168}$$

where the truncated approximation with n terms is given by

$$\widehat{f}_n(x;\alpha,\beta) = \frac{1}{\pi}\sum_{i=1}^{n}\frac{(-x)^{i-1}\Gamma(1+i/\alpha)}{i!r^i}\sin\left(\frac{i\pi}{2\alpha}(\alpha+s)\right) \tag{4.169}$$

for some tolerance level ε which could be set as an absolute or a relative truncation error. These results illustrate why this procedure should be done pointwise over the support $x > 0$ (i.e. at points at which you wish to evaluate the density). Next, we present the case of $\alpha = 1.6$ which is moderately heavy tailed with a range of β values for four grid points for x in Figure 4.10.

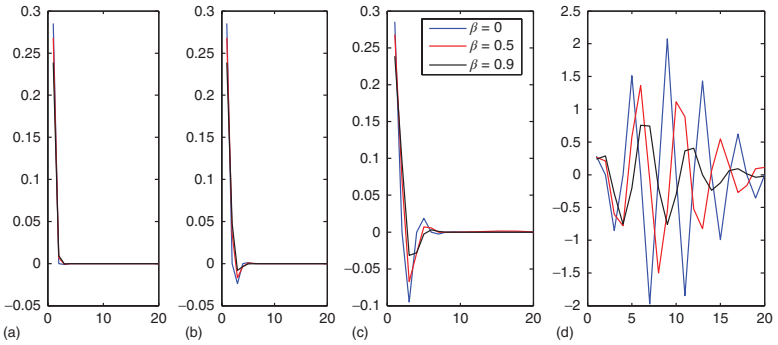

FIGURE 4.10 **Study of first 20 summand terms of the stable density series expansion for a range of parameter values at three different x locations. (a) $x = .1$, (b) $x = 0.5$, (c) $x = 1$ and (d) $x = 3$.** ***See insert for color representation of this figure.***

To complete this section on series representations for the family of α-stable severity models, we also consider series expansions that can be developed for the α-stable severity model via a Poisson process expansion, see details in Bentkus *et al.* (1996), Bentkus *et al.* (2001) and Samorodnitsky and Taqqu (1994a, section 1.4). In particular, one can show the following relationship between a Poisson process and a stable distribution detailed in Proposition 4.9.

Proposition 4.9 (Lévy LePage Series Representation) *Define the dependent sequence $\{\Gamma_i\}$ of arrival times from a Poisson process with unit rate; hence, each Γ_i is distributed as Gamma with shape i and mean of i. Furthermore, define the i.i.d. sequence of random variables $\{\varepsilon_i\}$ as a Bernoulli process defined by*

$$\varepsilon_i = \begin{cases} -1, & \Pr[\varepsilon_i = -1] = 0.5, \\ 1, & \Pr[\varepsilon_i = 1] = 0.5, \end{cases} \tag{4.170}$$

and the third i.i.d. sequence of random variables $\{W_i\}$ with a finite absolute αth moment. Consider these three sequences as independent from each other. Then given that $\alpha \in (0, 2)$, the sum given by

$$\lim_{n\to\infty} \sum_{i=1}^{n} \varepsilon_i \Gamma_i^{-1/\alpha} W_i \xrightarrow{d} X \sim S_\alpha \left(0, \left(C_\alpha^{-1} \mathbb{E}|W_1|^\alpha\right)^{1/\alpha}, 0; S1\right) \tag{4.171}$$

with constant term

$$C_\alpha = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, & \alpha \neq 1, \\ \dfrac{2}{\pi}, & \alpha = 1. \end{cases} \tag{4.172}$$

In a similar manner, one can also provide an infinite mixture of Gaussian representation for symmetric stable severity models. This representation is often used and is known as the scaled mixture of Gaussian representation, that is, the scale mixture of normals (SMiN) representation, see details in Godsill (2000b) and Peters *et al.* (2012d). In the case of the OpRisk severity model setting, we require to truncate such distributions to ensure that they have positive support as detailed next. This represents a practically relevant sub-family of α-stable distributions obtained when one considers restricting the skewness parameter β. A random variable X is said to be distributed from a symmetric α-stable distribution, $X \sim S_\alpha(0, \gamma, \delta; \cdot)$, when the skewness parameter $\beta = 0$. In this case, the model still captures a spectrum of distributions ranging from Gaussian $\alpha = 2$ through to infinite mean and infinite variance models. This particular sub-class is interesting as it can be represented uniquely by an SMiN representation as shown in Lemma 4.11.

Lemma 4.11 *In Andrews and Mallows (1974), West (1987) and Godsill (2000b, equations 3 and 4, p. 2), it is shown that when $X \sim S_\alpha(0, \gamma, \delta)$, it may be represented exactly by the following SMiN representation through the introduction of an auxiliary random variable λ,*

$$X|\lambda \sim Normal(\delta, \gamma\lambda), \tag{4.173}$$

with auxiliary scale variable $\lambda \sim S_{\alpha/2}(0, 1, 1; S1)$.

From the practical perspective in an OpRisk model, one is clearly interested in restriction of the support of this representation to take only positive values for the losses, such that $X \in \mathbb{R}^+$. So we are faced practically with the situation of the choice of either maintaining the convolutional properties we have discussed about α-stable models, at the cost of not obtaining simple representations for the density, or utilizing a simple representation such as the SMiN representation with a truncation to positive support, at the cost of losing the convolutional property. Both choices will be of relevance in applied OpRisk modeling. The truncated SMiN representation is given in Definition 4.27.

Definition 4.27 *A truncated SMiN α-stable random variable, denoted by $X \sim S_\alpha(x; 0, \gamma, \delta, S0)\mathbb{I}_{(0,\infty)}(x)$, may be represented exactly by the following SMiN representation through the introduction of an auxiliary random variable λ,*

$$X|\lambda \sim Z_\lambda^{-1} Normal(x; \delta, \gamma\lambda)\mathbb{I}_{(0,\infty)}(x), \tag{4.174}$$

with auxiliary scale variable $\lambda \sim S_{\alpha/2}(1,1,0)$*,* $X \in \mathbb{R}^+$ *and truncation normalizing probability for the conditional probability given by* $Z_\lambda = 1 - \Phi\left((0-\delta)/(\gamma\lambda)\right)$*. Therefore, the overall normalization for the marginal distribution* $X \sim S_\alpha(x;0,\gamma,\delta,S0)\mathbb{I}_{(0,\infty)}(x)$ *is given by*

$$Z = \int_0^\infty \left[1 - \Phi\left(\frac{0-\delta}{\gamma\lambda}\right)\right] dS_{\alpha/2}(1,1,0). \tag{4.175}$$

Remark 4.26 *The evaluation of the normalizing constant in the truncated SMiN representation can be approximated via an unbiased Monte Carlo estimate given by drawing* N *samples for the auxiliary variable exactly from the perfectly skewed stable distribution* $S_{\alpha/2}(1,1,0)$ *to obtain samples,* $\{\lambda_i\}_{i=1:N}$*. Then these samples are used to approximate the normalizing constant according to*

$$\widehat{Z} = \frac{1}{N}\sum_{i=1}^{N}\left[1 - \Phi\left(\frac{0-\delta}{\gamma\lambda_i}\right)\right]. \tag{4.176}$$

4.6 Distribution Representations of α-Stable Severity Models

In this section, we present the numerical evaluation of the distribution function of α-stable severity models. First, we consider the integral representations obtained for explicit representation of the distribution function of a standardized α-Stable severity model, which is the analogue for the density integral representation that was considered for the Zolotarev B-type parameterization in Theorem 4.18. The result is given in Theorem 4.19, see Zolotarev (1986, p. 78).

Theorem 4.19 (Integral Representation of the Distribution of Stable Severity Models) *The distribution function for the standard stable random variable with* $\gamma = 1$ *and* $\delta = 0$ *under the B-type parameterization can be evaluated using the following convergent series integral representation for* $x > 0$

$$F_X(x;\alpha,\beta;\ B) = \begin{cases} C(\alpha,\theta) + \frac{\varepsilon(\alpha)}{2}\int_{-\theta}^{1}\exp\left(-x^{\frac{1}{\alpha-1}}U_\alpha\left(\varphi,\theta\right)\right)\,d\varphi, & \alpha \neq 1, \\ \frac{1}{2}\int_{-1}^{1}\exp\left[-\exp\left(-\frac{x}{\beta}\right)U_1(\varphi,\beta)\right]\,d\varphi, & \alpha = 1 \text{ and } \beta > 0, \end{cases} \tag{4.177}$$

with the following functions given by

$$\begin{aligned} \varepsilon(\alpha) &= \operatorname{sgn}(1-\alpha),\ \ \theta^* = \theta\operatorname{sgn}(x),\ \ \theta = \beta\frac{K(\alpha)}{\alpha}, \\ C(\alpha,\theta) &= 1 - \frac{1}{4}(1+\theta)(1+\varepsilon(\alpha)), \\ U_\alpha(\varphi,\theta) &= \left(\frac{\sin\left(\frac{\pi}{2}\alpha(\varphi+\theta)\right)}{\cos\left(\frac{\pi}{2}\varphi\right)}\right)^{\frac{\alpha}{1-\alpha}}\frac{\cos\left(\frac{\pi}{2}\left((\alpha-1)\varphi+\alpha\theta\right)\right)}{\cos\left(\frac{\pi}{2}\varphi\right)}, \\ U_1(\varphi,\beta) &= \frac{\pi}{2}\frac{(1+\beta\varphi)}{\cos\left(\frac{\pi}{2}\varphi\right)}\exp\left(\frac{\pi}{2}\left(\varphi+\frac{1}{\beta}\right)\tan\left(\frac{\pi}{2}\varphi\right)\right). \end{aligned}$$

In addition, one can utilize the expression $F_X(-x;\alpha,\beta;\ B) + F_X(x;\alpha,-\beta;\ B) = 1$ *to evaluate the distribution for* $x < 0$*, and when* $x = 0$*, one has*

$$F_X(0;\alpha,\beta;\ B) = \frac{1}{2}\left(1 - \beta\frac{K(\alpha)}{\alpha}\right). \tag{4.178}$$

4.6.1 QUADRATURE APPROXIMATIONS FOR DISTRIBUTION REPRESENTATIONS OF α-STABLE SEVERITY MODELS

From this integral representation in Theorem 4.19, one can construct via one of the numerical integration approaches discussed previously an approximation of the distribution function pointwise. Consider the set of n points (for an n-point quadrature rule) given by $x_i \in [-\theta, 1]$ with $x_0 < x_1 < \cdots < x_n$ which produce the approximation

$$\widehat{F}_X(x;\alpha,\beta;\ B) = \begin{cases} C(\alpha,\theta) + \frac{1}{2}\varepsilon(\alpha)\sum_{j=0}^{n} a_j w(x_j)^{-1}\exp\left(-x^{1/(\alpha-1)}U_\alpha\left(x_j,\theta\right)\right), & \alpha \neq 1, \\ \frac{1}{2}\sum_{j=0}^{n} a_j w(x_j)^{-1}\exp\left[-\exp(-x/\beta)\,U_1(x_j,\beta)\right], & \alpha = 1 \text{ and } \beta > 0, \end{cases} \tag{4.179}$$

where $w(x)$ is a quadrature weight function (with orthogonal Askey polynomial pair of $q(x)$) such as the Gauss–Hermite choice given by $w(x) = \exp(-x^2)$. In the case of Gauss–Hermite quadratures, the weights are given at quadrature location x_j by

$$a_j = \frac{2^{n-1}n!\sqrt{\pi}}{n^2 H_{n-1}(x_j)^2}, \tag{4.180}$$

where $H_n(x)$ is the n-th order Hermite polynomial. The locations of the quadrature points in $[-\theta, 1]$ are given by finding the roots of the n-th order Hermite polynomial in this interval; this is typically achieved numerically via say the Newton method. To make such a procedure numerically robust, it is advisable to use the normalized n-th order Gauss–Hermite polynomials denoted by $\widetilde{H}_n(x)$ and given by the physicists definition

$$\widetilde{H}_n(x) = (-1)^n \exp(x^2)\frac{d^n}{dx^n}\exp(-x^2) \tag{4.181}$$

and the following recursive identities help with the application of the Newton method for root searching to find the required quadrature evaluation points

$$\begin{aligned} \widetilde{H}_{n+1}(x) &= \sqrt{\frac{2}{n+1}}x\widetilde{H}_n(x) - \sqrt{\frac{n}{n+1}}\widetilde{H}_{n-1}(x), \\ \frac{d}{dx}\widetilde{H}_n(x) &= \sqrt{2n}\widetilde{H}_{n-1}(x). \end{aligned} \tag{4.182}$$

The error in this approximation is then given for some value $\nu \in [-\theta, 1]$ by the expression, for example, in the case that $\alpha \neq 1$ and for any $x > 0$ by

$$\int_{-\theta}^{1} w(\varphi) \exp\left(-x^{\mu} U_{\alpha}(\varphi, \theta)\right) \, d\varphi - \sum_{j=0}^{n} a_j \exp\left(-x^{\mu} U_{\alpha}\left(x_j, \theta\right)\right)$$

$$= \frac{1}{(2n+2)!} \frac{d^{2n+2}}{d\varphi^{2n+2}} \left[\exp(-x^{\mu} U_{\alpha}(\varphi, \theta))\right]\Big|_{\varphi=\nu} \int_{-\theta}^{1} w(x) q_{n+1}(x)^2 dx, \tag{4.183}$$

where $\mu = 1/(\alpha - 1)$ and $q_{n+1}(x)$ is the $(n+1)$th order orthogonal polynomial corresponding to the Gauss quadrature weight, in this case, the Hermite polynomials (see discussion in Cruz *et al.* 2015).

4.6.2 CONVERGENT SERIES REPRESENTATIONS OF THE DISTRIBUTION FOR α-STABLE SEVERITY MODELS

In this section, we briefly present series representations for the distribution function which are analogous to those detailed previously for the density. These series expansions generalized earlier results developed by Humbert (1945), Pollard (1946) and later also by Zolotarev (1994). Under the Zolotarev type-B parameterization, one can also develop a closed-form series expansion for the α-stable distribution function in Theorem 4.20. This series representation is absolutely convergent if $\alpha \in (1, 2]$ for the range $x > 0$ and if $\alpha \in (0, 1)$; it acts as an asymptotic expansion of the α-stable distribution as $x \to 0$, see details in Zolotarev (1986, p. 94–95).

Theorem 4.20 (Representation 1 of the Distribution Function of Stable Severity Models) *The distribution function for the standard stable random variable with $\gamma = 1$ and $\delta = 0$ under the B-type parameterization can be evaluated using the following convergent series integral representation for $x > 0$*

$$F_X(x; \alpha, \beta, 0, 1; B) = \sum_{n=0}^{\infty} \frac{1}{n!} f_n x^n, \tag{4.184}$$

where the coefficients are given by

$$f_n = \begin{cases} \dfrac{1}{2}\left(1 - \dfrac{\beta K(\alpha)}{\alpha}\right), & n = 0, \\ (-1)^{n-1} \dfrac{1}{\pi} \dfrac{\Gamma(1 + n/\alpha)}{n} \sin\left(\dfrac{\pi n}{2}\left(1 + \dfrac{\beta K(\alpha)}{\alpha}\right)\right), & n \geq 1 \end{cases} \tag{4.185}$$

with the relationship for $x \leq 0$ given by

$$F_X(x; \alpha, \beta, 0, 1; B) = 1 - F_X(-x; \alpha, -\beta, 0, 1; B). \tag{4.186}$$

One can also develop a second series representation in terms of fractional polynomial powers given in Theorem 4.21. This series representation is absolutely convergent if $\alpha \in (0, 1)$ for the range $x > 0$ and if $\alpha \in (1, 2]$; it acts as an asymptotic expansion of the α-stable distribution as $x \to \infty$, see details in Zolotarev (1986, p. 94–95).

Theorem 4.21 (Representation 2 of the Distribution Function of Stable Severity Models) *The distribution function for the standard stable random variable with $\gamma = 1$ and $\delta = 0$ under the B-type parameterization can be evaluated using the following convergent series integral representation for $x > 0$*

$$F_X(x;\alpha,\beta,0,1;B) = \sum_{n=0}^{\infty} \frac{1}{n!}\widetilde{f}_n x^{-\alpha n}, \tag{4.187}$$

where the coefficients are given by

$$\widetilde{f}_n = \begin{cases} 1, & n = 0 \\ (-1)^{n-1}\dfrac{1}{\alpha\pi}\dfrac{\Gamma(\alpha n+1)}{n}\sin\left(\dfrac{\pi n}{2}(\alpha+\beta K(\alpha))\right), & n \geq 1 \end{cases} \tag{4.188}$$

with the relationship for $x \leq 0$ given by

$$F_X(x;\alpha,\beta,0,1;B) = 1 - F_X(-x;\alpha,-\beta,0,1;B). \tag{4.189}$$

The properties of this expansion are that for $\alpha \in [1,2]$, this series expansion diverges. However, the expansion of the first n terms (for all n) has an error of known order of magnitude for a given series truncation according to

$$F(x;\alpha,\beta;B) = \left(\sum_{k=0}^{n} \frac{1}{k!}\widetilde{f}_k x^{-\alpha k} + \beta - \alpha\arg(x)\right) + O(|x|^{-\alpha(n+1)-1}). \tag{4.190}$$

4.7 Quantile Function Representations and Loss Simulation for α-Stable Severity Models

In this section, we present two important aspects of the α-stable severity model that practitioners utilize. In particular, when calculating tail functionals and risk measures such as value-at-risk (VaR) and expected shortfall. The first involves the numerical evaluation of the quantile function of the α-stable severity model. The approach adopted below will be based on a series expansion representation that utilizes Lagrange's inversion formula and is studied for the asymmetric α-stable model case in Munir and Shaw (2012). In addition, we will then show how to sample exactly stable loss random variables which will involve a simple rejection algorithm first developed for general characteristic functions by Devroye (1981, 1984) and then developed for the specific case of α-stable models by Chambers *et al.* (1976); also see Weron (1996b, 1996a).

4.7.1 APPROXIMATING THE QUANTILE FUNCTION OF STABLE LOSS RANDOM VARIABLES

Here we consider a loss random variable $X \sim S(\alpha,\beta,\gamma,\delta;1)$ under Nolan's $S1$ parameterization, and we define the quantile function by $Q_X(\alpha) = F_X^{-1}(u)$ where $\mathbb{P}\mathrm{r}[X \leq x_u] = u$ for some quantile loss level denoted by x_u and probability $u \in [0,1]$. If the density function $f_X(x)$ for the loss random variable was known, then one could resort to differentiating the quantile relationship to obtain the representation

$$\frac{dQ_X(u)}{du} = \frac{1}{F_X'(F_X^{-1}(u))} = \frac{1}{f_X(Q_X(u))}. \tag{4.191}$$

As noted in Munir and Shaw (2012), the non-linear terms in this equation occur because of the density function, which when the density has a complicated representation, can result in significant challenge in finding any type of closed-form representation for the quantile function. Finding the numerical solution can also be challenging, although it is argued in Munir and Shaw (2012) that if one were to consider instead the numerical solution or series representation for the solution, this will be possible if the inverse of the density $1/f_X(x)$ is an analytic function at the initial value $x_0 = Q_X(u_0)$.

Definition 4.28 (Analytic Function) *An analytic function is any function that can be locally expressed by a convergent power series,*

$$f(x) = \sum_{n=0}^{\infty} a_n (x - x_0)^n. \tag{4.192}$$

It is, therefore, also true that a function is analytic if and only if its Taylor series about x_0 converges to the function in some neighbourhood for every x_0 in its domain. As has been demonstrated above, one can find a convergent power series representation for the α-stable density. Hence, one may adopt a Taylor series expansion to represent the quantile function solution to Equation 4.191.

When evaluating the quantile function near its singular points, this approach must be done with caution as a straight series expansion will result in numerical difficulties when solving for the coefficients, because the system of equations that results can be ill-conditioned. In Munir and Shaw (2012), this ill-conditioning numerical challenge is overcome by introduction of a change of variable as discussed in Shaw *et al.* (2011). The proposed change of variable $z = Q_Y(U)$ is analogous in nature to a careful choice of the base distribution when representing quantile functions of a desired distribution as transformations of, the so-called, 'base distributions'. In this case, the base distribution denoted by its quantile function for the base random variable Y is given by Q_Y and defines the transformation. This concept goes back a lot longer than the paper of Shaw *et al.* (2011) and indeed the ideas of this form can be traced a long way back in the statistics literature, see, for example, the detailed discussion on this point in Cruz *et al.* (2015) where the Tukey family of quantile function transformations are considered. It was also used to develop generalizations of the Cornish Fisher density expansions.

Hence, by introducing the change of variable with a new base distribution, one obtains the following representation for the target quantile function $Q_X(u)$ given in terms of the transformation $T(\cdot)$ and the based quantile function $Q_Y(u)$ according to $Q_X(u) = T(Q_Y(u))$. Then by looking at this with regard to Equation 4.191, one obtains, under the transformation $T(z)$ where the change of base quantile function involves $z = Q_Y(u)$, the expression derived via the chain rule as follows:

$$\begin{aligned}
\frac{dQ_X(u)}{du} &= \frac{dT}{dz}\frac{dz}{du} \\
&= \frac{dT}{dz}\frac{dQ_Y(u)}{du} \\
&= \frac{dT}{dz}\frac{1}{f_Y(Q_Y(u))} \\
&= \frac{dT}{dz}\frac{1}{f_Y(z)},
\end{aligned} \tag{4.193}$$

where the term $\frac{dT}{dz}$ is denoted as the first-order recycling condition and is given by

$$\frac{dT}{dz} = \frac{f_Y(z)}{f_X(T(z))}. \tag{4.194}$$

In the α-stable case, the approach adopted in Shaw *et al.* (2011) is to proceed via the Lagrange inversion formula. The Lagrange inversion formula given in Theorem 4.22, see Lagrange (1770), is a classical result which had a significant influence on several areas of mathematics. This inversion theorem was later extended in the work of Bürmann (1796).

Theorem 4.22 (Lagrange Inversion Formula) *Consider the functional relationship $f(x) = y$ where the function f is analytic at point a and the derivative at this point is non-zero $f'(a) \neq 0$, then the solution to the inverse function, given by $f^{-1}(y) = x$, in the neighbourhood of $f(a)$, with the function f^{-1} analytic at the point $f(a)$ is given by the series expansion*

$$f^{-1}(y) = a + \sum_{n=1}^{\infty} \left[\lim_{x \to a} \left(\frac{(y - f(a))^n}{n!} \frac{d^{n-1}}{dx^{n-1}} \left(\frac{x - a}{f(x) - f(a)} \right)^n \right) \right]. \tag{4.195}$$

The solution obtained numerically for the α-stable family by Shaw *et al.* (2011) then proceeds as follows by first considering the generic Taylor series representations of the α-stable distribution function and quantile functions:

$$\begin{aligned} F_X(x) &= \sum_{n=1}^{\infty} \frac{1}{n} f_n (x - x_0)^n, \\ Q_X(u) &= \sum_{n=1}^{\infty} \frac{1}{n} q_n (u - u_0)^n \end{aligned} \tag{4.196}$$

with the relationship discussed that $F_X(Q_X(u)) = u$ for all $u \in (0, 1)$. Then given a convergent series expansion representation of the distribution for the α-stable model (see Theorems 4.20 and 4.21), the challenge will be to utilize the Lagrange inversion theorem to find an expression for the quantile function coefficients in terms of the known coefficients for the distribution function expansion. The Lagrange inversion theorem then states that one can obtain the expression, which was further refined in the solution provided by Aldrovandi (2001) to produce

$$\begin{aligned} q_n &= \begin{cases} (f_1)^{-1}, & n = 1, \\ -(f_1)^{-n} \sum\limits_{k=1}^{n} q_k B_{n,k}\left(f_1, f_2, \ldots, f_{n-k+1}\right), & n \geq 2, \end{cases} \\ &= \begin{cases} (f_1)^{-1}, & n = 1, \\ -(f_1)^{-n} \sum\limits_{k=1}^{n} (-1)^k \frac{(n+k-1)!}{(n-1)!} B_{n-1,k}\left(\frac{f_2}{2f_1}, \frac{f_3}{3f_1}, \ldots, \frac{f_{n-k+1}}{(n-k+1)f_1}\right), & n \geq 2, \end{cases} \end{aligned} \tag{4.197}$$

with $B_{n,k}(f_1, f_2, \ldots, f_{n-k+1})$ denoting the Bell polynomials given by

$$\begin{aligned} &B_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) \\ &\quad = \sum \frac{n!}{j_1! j_2! \cdots j_{n-k+1}!} \left(\frac{x_1}{1!}\right)^{j_1} \left(\frac{x_2}{2!}\right)^{j_2} \cdots \left(\frac{x_{n-k+1}}{(n-k+1)!}\right)^{j_{n-k+1}} \end{aligned} \tag{4.198}$$

with the summation taken over all sequences $j_1, j_2, \ldots, j_{n-k+1}$ of non-negative integers that satisfy $j_1 + j_2 + \cdots = k$ and $j_1 + 2j_2 + 3j_3 + \cdots = n$.

Using this inversion result and the series expansions, which have rapid convergence properties for the central and tail regions given, respectively, by Theorems 4.20 and 4.21, one can obtain the quantile expansions $Q_X(u)$ for $u > u_0$ where u_0 is the zero quantile given by $u_0 = F_X(0; \alpha, \beta, 0, 1; B)$. This series representation derived for the quantile function is absolutely convergent if $\alpha \in (1, 2]$ for the range $u > u_0$ and if $\alpha \in (0, 1)$; it acts as an asymptotic expansion of the α-stable distribution as $u \to u_0$, see Theorem 4.23.

Theorem 4.23 (Representation 1 of the Quantile Function of Stable Severity Models) *The quantile function for the standard stable random variable with $\gamma = 1$ and $\delta = 0$ under the B-type parameterization can be evaluated using the following convergent series integral representation for $u > u_0$*

$$Q_X(u; \alpha, \beta, 0, 1; B) = \sum_{n=0}^{\infty} \frac{1}{n!} q_n (u - u_0)^n \tag{4.199}$$

with $u_0 = F_X(0; \alpha, \beta, 0, 1; B) = \frac{1}{2}(1 - \beta K(\alpha)/\alpha)$ where the coefficients are given by

$$q_n = \begin{cases} (f_1)^{-1}, & n = 1, \\ -(f_1)^{-n} \sum_{k=1}^{n} (-1)^k \frac{(n+k-1)!}{(n-1)!} B_{n-1,k}\left(\frac{f_2}{2f_1}, \frac{f_3}{3f_1}, \ldots, \frac{f_{n-k+1}}{(n-k+1)f_1}\right), & n \geq 2, \end{cases} \tag{4.200}$$

with

$$f_n = \begin{cases} \frac{1}{2}(1 - \beta K(\alpha)/\alpha), & n = 0, \\ (-1)^{n-1} \frac{1}{\pi} \frac{\Gamma(1+n/\alpha)}{n} \sin\left(\frac{\pi n}{2}(1 + \beta K(\alpha)/\alpha)\right), & n \geq 1, \end{cases} \tag{4.201}$$

and the relationship for $u \leq u_0$ is then given by

$$Q_X(u; \alpha, \beta, 0, 1; B) = -Q_X(1 - u; \alpha, -\beta, 0, 1; B). \tag{4.202}$$

As with the distribution function series expansions, one can also develop a second series representation in terms of fractional polynomial powers given in Theorem 4.24. This series representation is absolutely convergent if $\alpha \in (0, 1)$ for the range $u > u_0$ and if $\alpha \in (1, 2]$; it acts as an asymptotic expansion of the α-Stable distribution as $u \to 1$. The representation in Theorem 4.24 is obtained in Shaw *et al.* (2011) by making a change of variable $y = x^{1/\alpha}$ and then applying Lagrange's inversion formula to the transformed power series given in Theorem 4.21.

Theorem 4.24 (Representation 2 of the Quantile Function of Stable Severity Models) *The quantile function for the standard stable random variable with $\gamma = 1$ and $\delta = 0$ under the B-type parameterization can be evaluated using the following convergent series integral representation for $u > u_0$*

$$Q_X(u; \alpha, \beta, 0, 1; B) = \left[\sum_{n=0}^{\infty} \frac{1}{n!} \widetilde{q}_n (u - 1)^n\right]^{1/\alpha} \tag{4.203}$$

with $u_0 = F_X(0;\alpha,\beta,0,1;B) = \frac{1}{2}\left(1 - \frac{\beta K(\alpha)}{\alpha}\right)$ *where the coefficients are given by*

$$\widetilde{q}_n = \begin{cases} (\widetilde{f}_1)^{-1}, & n = 1, \\ -(\widetilde{f}_1)^{-n} \sum_{k=1}^{n} (-1)^k \frac{(n+k-1)!}{(n-1)!} B_{n-1,k}\left(\frac{\widetilde{f}_2}{2\widetilde{f}_1}, \frac{\widetilde{f}_3}{3\widetilde{f}_1}, \ldots, \frac{\widetilde{f}_{n-k+1}}{(n-k+1)\widetilde{f}_1}\right), & n \geq 2, \end{cases} \tag{4.204}$$

with

$$\widetilde{f}_n = \begin{cases} 1, & n = 0, \\ (-1)^{n-1}(\alpha\pi n)^{-1}\Gamma(\alpha n + 1)\sin\left(\frac{\pi n}{2}(\alpha + \beta K(\alpha))\right), & n \geq 1, \end{cases} \tag{4.205}$$

and the relationship for $u \leq u_0$ *is then given by*

$$Q_X(u;\alpha,\beta,0,1;B) = -Q_X(1-u;\alpha,-\beta,0,1;B). \tag{4.206}$$

4.7.2 SAMPLING REALIZATIONS OF STABLE LOSS RANDOM VARIABLES

The following provides an exact and efficient algorithm for sampling via transformation from Nolan's $S1$ parameterization of the α-stable severity model, see further details in Nolan (2013) and the original formulation in Chambers (1976).

Algorithm 4.1 Simulating Losses from Nolan's $S1$-Parameterized α-Stable Severity Models

1. *Draw a uniform random variable* $U \sim$ *Uniform* $(-\pi/2, \pi/2)$.
2. *Draw an exponential random variable* $E \sim Exp(1)$ *independently of* U.
3. *If* $\alpha \neq 1$, *then evaluate functions*

$$\begin{aligned} B_{\alpha,\beta} &= \frac{\arctan\left(\beta\tan\left(\pi\alpha/2\right)\right)}{\alpha}, \\ S_{\alpha,\beta} &= \left(1 + \beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{1/(2\alpha)}. \end{aligned} \tag{4.207}$$

4. *Evaluate the transformation*

$$Z = \begin{cases} S_{\alpha,\beta}\dfrac{\sin\left(\alpha\left(U + B_{\alpha,\beta}\right)\right)}{\cos^{1/\alpha}(U)}\left[\dfrac{\cos\left(U - \alpha\left(U + B_{\alpha,\beta}\right)\right)}{E}\right]^{(1-\alpha)/\alpha}, & \alpha \neq 1 \\ \dfrac{2}{\pi}\left[\left(\dfrac{\pi}{2} + \beta U\right)\tan U - \beta\ln\left(\dfrac{\frac{\pi}{2}E\cos U}{\frac{\pi}{2} + \beta U}\right)\right], & \alpha = 1, \end{cases} \tag{4.208}$$

   where $Z \sim S(\alpha,\beta,0,1;\ S1)$ *where* $B_{\alpha,\beta}$ *accounts for a skewness parameter change and* $S_{\alpha,\beta}$ *accounts for a scale parameter change.*
5. *To obtain a loss random variable* $X \sim S(\alpha,\beta,\gamma,\delta;\ S1)$, *apply the transformation*

$$X = \begin{cases} \gamma Z + \delta, & \alpha \neq 1, \\ \gamma Z + \frac{\pi}{2}\beta\gamma\ln\gamma + \delta, & \alpha = 1. \end{cases} \tag{4.209}$$

4.8 Parameter Estimation in an α-Stable Severity Model

Many authors have studied the different approaches one may adopt in performing parameter estimation for the α-stable family of severity models, given observed loss data. A summary of the techniques that have been considered includes the following.

1. *Tail Index of Stability Estimators.* Tail estimators based on regression of the empirical distribution function or the empirical characteristic function; see discussions in Blattberg and Sargent (1971), McCulloch (1998), Fofack and Nolan (1999), Marinelli and Rachev (2000) and the critique of such approaches in DuMouchel (1983). In this class of estimators for the tail index, there are also the variants considered based on the Hill estimator, see discussions in Daniélsson *et al.* (2001) and Daniélsson and De Vries (1997). There have also been extensions to these classes of estimators where people have tried the Pickands estimator, see discussion in Richards *et al.* (2012), and the modified unconditional Pickands estimator of Mittnik and Rachev (1991). For estimation of all the parameters of a stable law under a regression-based procedure, see the approach of Koutrouvelis (1980) and the refined version in Koutrouvelis (1981).
2. *Quantile-Based Parameter Estimators.* In the class of parameter estimators for the α-stable severity model, the quantile-based estimators have proven to be very successful and widely used. These include the works of McCulloch (1986) and the subsequent modification of Marinelli and Rachev (2000).
3. *Characteristic Function Estimators.* In the class of characteristic-function-based methods, there are studies such as those of Zolotarev (1986) who works under Koutrouvelis (1982) and improvements in Besbeas and Morgan (2008). In general, the properties of such procedures, which are based on use of the characteristic function are studied in Feuerverger and McDunnough (1981) and Yu (2004).
4. *Maximum Likelihood Estimators.* In the class of maximum likelihood estimation (MLE) procedures, there is the highly successful approach of Nolan (2001) and Mittnik *et al.* (1999b).
5. *Generalized Moment-Matching Estimators.* The most well known of these methods is the approach proposed by Press (1972).
6. *Indirect Inference Estimators.* In the class of indirect inference methods for estimation of stable parameters, there is the work of Garcia *et al.* (2006).
7. *Bayesian Models.* In the class of Bayesian models, there have been numerous approaches to estimation, including the auxiliary variable Markov chain Monte Carlo (MCMC) methods of Godsill (1999) which developed a combination of the expectation maximization and MCMC algorithms, the work of Godsill (2000b) where a slice sampler was considered and the auxiliary variable and transformation MCMC approach of Buckle (1995) and the simple MCMC- and FFT-based approach of Casarin (2003) with a series expansion for the tails. In the class of approximate Bayesian computation (ABC) methods, there were also SMC solutions developed in Peters *et al.* (2012b).
8. *Wavlet-Based Estimators.* There have not been many solutions developed for wavelet methods, see Antoniadis *et al.* (2006).

There have also been some comparative studies performed for different estimation procedures, see, for instance, Akgiray and Lamoureux (1989). In the following, we provide some basic details for some of the procedures that are numerically robust and relatively straightforward to implement.

4.8.1 McCULLOCH'S QUANTILE-BASED α-STABLE SEVERITY MODEL ESTIMATORS

The method used to estimate the parameters for the α-stable is based on the quantile approach of McCulloch (1998). This approach was selected because it is known to be robust and computationally efficient. Estimates of the model parameters are based on sample quantiles, while correcting for estimator skewness because of the evaluation of the sample quantiles $\widehat{q}_p(x)$, where $\widehat{q}_p(x)$ denotes the pth quantile of x from the order statistics of the sample. The four parameters in the α-stable model under the Nolan $S1$ parameterization presented were determined from a set of five pre-determined quantiles for the parameter ranges $\alpha \in [0, 2.0]$, $\beta \in [-1, 1]$, $\gamma \in [0, \infty)$ and $\delta \in \mathbb{R}$ as detailed in McCulloch (1986).

The approach uses differences between quantiles, properly normalized, to break dependence on location and scale parameters. Then two functions of the stability index α and the skewness β are numerically calculated via the sample quantile values and then numerically inverted to obtain the parameter estimators. The use of an interpolation method is considered to improve the accuracy.

The stages of this estimation involve the following details:

1. *Obtain a Finite Sample Consistent Estimator of Quantiles.* With the $x_{(i,n)}$ arranged in ascending order, the skewness correction is made by matching the sample order statistics with $\widehat{q}_{s(i)}(x)$ where $s(i) = (2i-1)/(2n)$. Then a linear interpolation to p from the two adjacent $s(i)$ values is used to establish $\widehat{q}_p(x)$ as a consistent estimator of the true quantiles. This corrects for spurious skewness present in finite samples and $\widehat{q}_p(x)$ is a consistent estimator of $q_p(x)$. Here, one would have for a sample of size n with $s(i) < p \leq s(i+1)$ the linearly interpolated pth quantile has corrected value

$$\widehat{q}_p(x) = n(\widehat{q}_{s(i+1)}(x) - \widehat{q}_{s(i)}(x))(p - \widehat{q}_{s(i)}(x)). \tag{4.210}$$

2. *Obtain Estimates of Tail Index α and Skewness Parameter β.* In McCulloch (1986), two non-linear functions of α and β are provided in terms of the quantiles as detailed in Equation 4.211.

$$\nu_\alpha = \frac{q_{0.95}(\cdot) - q_{0.05}(\cdot)}{q_{0.75}(\cdot) - q_{0.25}(\cdot)},\ \nu_\beta = \frac{q_{0.95}(\cdot) + q_{0.05}(\cdot) - 2q_{0.5}(\cdot)}{q_{0.95}(\cdot) - q_{0.05}(\cdot)}. \tag{4.211}$$

   We can, therefore, estimate these quantities $\widehat{\nu}_\alpha$ and $\widehat{\nu}_\beta$ using the sample estimates of the quantiles $\widehat{q}_p(x)$. Now to obtain the actual parameters estimates $\widehat{\alpha}$ and $\widehat{\beta}$, we numerically invert the non-linear functions ν_α and ν_β. This can be done efficiently through a look-up table provided for numerous combinations of α and β and provided in tabulated form in McCulloch (1986). This approach yields consistent estimators for $\widehat{\alpha}$ and $\widehat{\beta}$.

3. *Obtain Estimates of γ Given Estimates of α and β.* In McCulloch (1986), a third non-linear function which is explicit in γ and implicit in α and β is provided in terms of the quantiles

as detailed in Equation 4.212.

$$\nu_{\gamma}(\alpha,\beta)=\frac{q_{0.75}(\cdot)-q_{0.25}(\cdot)}{\gamma}, \tag{4.212}$$

an estimate then follows given $\widehat{\alpha}$, $\widehat{\beta}$ and consistent sample quantiles $\widehat{q}_{0.75}, \widehat{q}_{0.25}$. This yields the following consistent estimator of γ given by

$$\widehat{\gamma}=\frac{q_{0.75}(\cdot)-q_{0.25}(\cdot)}{\nu_{\gamma}(\widehat{\alpha},\widehat{\beta})}. \tag{4.213}$$

The function $\nu_{\gamma}(\widehat{\alpha},\widehat{\beta})$ is then estimated by a bicubic discretized version which is then utilized to make the estimation.

4. *Obtain Estimates of δ Given Estimates of α, β and γ.* In McCulloch (1986), a new parameter is utilized as given by

$$\zeta=\begin{cases}\delta+\beta\gamma\tan\left(\pi\alpha/2\right), & \alpha\neq 1,\\ \delta, & \alpha=1.\end{cases} \tag{4.214}$$

Then one defines the function

$$\nu_{\zeta}(\alpha,\beta)=\frac{\zeta-q_{0.5}(\cdot)}{\gamma}, \tag{4.215}$$

which gives the estimator

$$\widehat{\zeta}=\widehat{q}_{0.5}(\cdot)+\widehat{\gamma}\nu_{\zeta}(\widehat{\alpha},\widehat{\beta}). \tag{4.216}$$

This in turn yields the estimator after transformation back to the B-type parameterization given by

$$\widehat{\delta}=\widehat{\zeta}-\widehat{\beta}\widehat{\gamma}\tan\left(\frac{\pi\widehat{\alpha}}{2}\right). \tag{4.217}$$

See discussions on this approach in McCulloch (1986) and the subsequent modification of Marinelli and Rachev (2000).

4.8.2 ZOLOTAREV'S TRANSFORMATION TO W-CLASS-BASED α-STABLE SEVERITY MODEL ESTIMATORS

On the basis of a transformation of data from the α-stable family $X \to Z$, Zolotarev (1986, p. 16 and Chapter 4) provides an alternative parameterization of the α-stable model $(\alpha,\beta,\gamma,\delta)\leftrightarrow(\nu,\eta,\tau)$ with a characteristic function of the form (denoted as type-W) and given by

$$\ln\Phi_{Z}(\theta)=-\exp\left\{\nu^{-\frac{1}{2}}\left[\ln|\theta|+\tau-\mathrm{i}\frac{\pi}{2}\eta\,\mathrm{sgn}(\theta)\right]+\mathbb{C}\left(\nu^{-\frac{1}{2}}-1\right)\right\}, \tag{4.218}$$

where $\mathbb{C}$ is Euler's constant and $\nu\geq 1/4$, $|\eta|\leq\min\{1,2\sqrt{\nu}-1\}$ and $|\tau|<\infty$.

The transformation from the type-B parameterization of the stable characteristic function to the type-W above is achieved by the relationships

$$\nu = \alpha^{-2},$$

$$\theta = \begin{cases} \beta K(\alpha)/\alpha, & \alpha \neq 1, \\ \frac{2}{\pi} \arctan\left(\frac{2}{\pi}\delta\right), & \alpha = 1, \end{cases}$$

$$\tau = \begin{cases} \alpha^{-1} \ln \gamma + \mathbb{C}\left(\alpha^{-1} - 1\right), & \alpha \neq 1, \\ \ln \gamma + \frac{1}{2} \ln\left(\delta^2 + \pi^2/4\right), & \alpha = 1, \end{cases}$$

and the inverse relationships given by

$$\alpha = \frac{1}{\sqrt{\nu}},$$

$$\beta = \theta \max\left\{1, \frac{1}{2\sqrt{\nu} - 1}\right\} \operatorname{sgn}\left(1 - 1/\nu\right),$$

$$\gamma = \begin{cases} \exp(\tau/\sqrt{\nu} - \mathbb{C}\left(1 - 1/\sqrt{\nu}\right)), & \nu \neq 1, \\ \exp(\tau + (\ln \cos\left(\pi\theta/2\right) - \ln(\pi/2))), & \nu = 1, \end{cases}$$

$$\delta = \begin{cases} 0, & \nu \neq 1, \\ \frac{\pi}{2} \tan\left(\pi\theta/2\right), & \nu = 1. \end{cases}$$

This parameterization has the advantage that logarithmic moments have simple expressions in terms of parameters to be estimated. For a fixed constant $0 < \xi \leq 1/2$ (Zolotarev (1986) recommends $\xi = 0.25$) and for integer $n/3$, the transformation is

$$Z_j = X_{3j-2} - \xi X_{3j-1} - (1 - \xi) X_{3j}, \qquad j = 1, 2, \ldots, n/3.$$

Defining $V_j = \ln|Z_j|$ and $U_j = \operatorname{sgn}(X_j)$, estimates for ν, η and τ are then given by

$$\widehat{\nu} = \max\{\widetilde{\nu}, (1 + |\widehat{\eta}|)^2/4\}, \quad \widehat{\eta} = \mathbb{E}[U], \quad \widehat{\tau} = \mathbb{E}[V],$$

where

$$\widetilde{\nu} = \frac{6}{\pi^2} S^2(V) - \frac{3}{2} S^2(U) + 1,$$

using sample variances $S^2(V)$ and $S^2(U)$. As before, δ is estimated by $\frac{1}{n}\sum_{i=1}{}^{n} X_i$ (for $\alpha > 1$). Hence, one can then use the inverse parameter transforms to find the estimators for the type-B parameterization.

4.8.3 PRESS'S METHOD-OF-MOMENTS-BASED α-STABLE SEVERITY MODEL ESTIMATORS

For $\alpha \neq 1$ and unique evaluation points t_1, t_2, t_3, t_4, the method of moment's equations obtained from $\ln \Phi_X(t)$ can be solved to obtain (Press 1972; Weron 2006)

$$\ln(\widehat{\gamma}) = \frac{\ln|t_1| \ln(-\ln|\Phi(t_2)|) - \ln|t_2| \ln(-\ln|\Phi(t_1)|)}{\ln|t_1/t_2|}, \quad \widehat{\alpha} = \frac{\ln \frac{\ln|\Phi(t_1)|}{\ln|\Phi(t_2)|}}{\ln|t_1/t_2|},$$

$$\widehat{\beta}|_{\widehat{\alpha},\widehat{\gamma}} = \frac{\widehat{u}(t_4)/t_4 - \widehat{u}(t_3)/t_3}{(|t_4|^{\widehat{\alpha}-1} - |t_3|^{\widehat{\alpha}-1})\widehat{\gamma}^{\widehat{\alpha}} \tan\left(\frac{\widehat{\alpha}\pi}{2}\right)}, \; \widehat{\delta}|_{\widehat{\alpha}} = \frac{|t_4|^{\widehat{\alpha}-1}\widehat{u}(t_3)/t_3 - |t_3|^{\widehat{\alpha}-1}\widehat{u}(t_4)/t_4}{|t_4|^{\widehat{\alpha}-1} - |t_3|^{\widehat{\alpha}-1}},$$

where

$$\widehat{u}(t) = \tan^{-1}\left[\sum_{i=1}^{n}\cos(tx_i)/\sum_{i=1}^{n}\sin(tx_i)\right].$$

We adopt the evaluation points $t_1 = 0.2, t_2 = 0.8, t_3 = 0.1$ and $t_4 = 0.4$ as recommended by Koutrouvelis (1980) and accordingly obtain estimators for the parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\gamma}, \widehat{\delta})$.

4.9 Location of the Most Probable Loss Amount for Stable Severity Models

The most probable loss amount under an α-stable severity model is the focus of this section, and in particular, how it varies as a function of tail index and skewness parameter. Although it is known that all α-stable densities are unimodal, it, however, remains that there is no known formula for the location of the mode, denoted by $m(\alpha, \beta; k)$, in the $S0$ and $S1$ parameterizations of Nolan corresponding to $k \in \{0, 1\}$. The two exceptions to this include the symmetric $S0$ parameterization in which the mode $m(\alpha, 0; 0) = 0$, see Wintner (1936) and Fofack and Nolan (1999), and the parametric case of the perfectly skewed Lévy distribution which is of particular interest in OpRisk, see Peters *et al.* (2011). In the case of the Lévy distribution of relevance to OpRisk with strictly positive support, one has $m(1/2, -1; 0) = 2/3$.

In all other cases, it has been observed numerically and discussed in detail in Nolan (2015) that under the $S0$ parameterization the following results are observed in relation to the mass either side of the mode for a random variable $X \sim S(\alpha, \beta, 1, 0; S0)$ one has the proportion of mass either side of the mode satisfying $\mathbb{P}r[X > m(\alpha, \beta; 0)] > \mathbb{P}r[X < m(\alpha, \beta; 0)]$ for the $\beta > 0$; in the symmetric case, $\mathbb{P}r[X > m(\alpha, \beta; 0)] = \mathbb{P}r[X < m(\alpha, \beta; 0)]$; and in the case of $\beta < 0$, one has $\mathbb{P}r[X > m(\alpha, \beta; 0)] < \mathbb{P}r[X < m(\alpha, \beta; 0)]$.

In addition, it is shown in Fofack and Nolan (1999) that the mode of the $S1$ parameterization is defined with respect to the $S0$ parameterization model according to

$$m(\alpha, \beta; 1) = m(\alpha, \beta; 0) + \beta \tan\left(\frac{\pi\alpha}{2}\right). \tag{4.219}$$

Remark 4.27 *Furthermore, it is known that all non-degenerate stable distributions are continuous distributions with an infinitely differentiable density, see Nolan (2015, theorem 1.9)*

This results is particularly used for numerical procedures utilized to obtain estimates of the mode of the α-stable distribution. In addition, it can be used, as in Fofack and Nolan (1999), to study the rate at which the mode goes to zero: $\lim_{\beta\to 0} m(\alpha, \beta; 0) \to 0$. This rate can be defined by the partial derivative, which is a function of the tail index and a limit on the skewness parameter according to

$$R(\alpha) = \lim_{\beta\to 0}\frac{m(\alpha, \beta; 0)}{\beta} = \left.\frac{\partial m(\alpha, \beta; 0)}{\partial\beta}\right|_{\beta=0}. \tag{4.220}$$

In Fofack and Nolan (1999), this rate is derived and given in Lemma 4.12 as a function of the tail index.

Lemma 4.12 *Given $R(\alpha)$ is defined according to Equation 4.220 as the rate at which the mode under the $S0$ parameterization approaches 0 as the skewness parameter β goes to 0, then it is given analytically by*

$$R(\alpha) = \begin{cases} \tan(\pi\alpha/2)\left[\frac{\Gamma(1+2/\alpha)}{\Gamma(3/\alpha)} - 1\right], & \alpha \neq 1, \\ (2C_\gamma - 3)/\pi, & \alpha = 1, \end{cases} \tag{4.221}$$

where $C_\gamma \approx 0.57721$ is Euler's constant.

The proof of this result is based on modification of a result in Hall (1984, theorem 2.1) and Gawronski and Wiessner (1992).

Remark 4.28 *As the $S0$ parameterization is considered, the rate of convergence of the mode $m(\alpha, \beta; 0) \to 0$ as $\beta \to 0$ is well defined for $\alpha = 1$ and can be given by*

$$\lim_{\alpha \to 1} R(\alpha) = \frac{2C_\gamma - 3}{\pi}. \tag{4.222}$$

In addition, it can be shown that the distribution of the standardized stable distribution will be bounded by the symmetric stable distribution at its mode, as derived in Fofack and Nolan (1999) and given in Lemma 4.13.

Lemma 4.13 *Defining the density function of a standardized stable distribution by $f(x; \alpha, \beta, 1, 0; S0)$, for all $x \in \mathbb{R}$ and $\beta \in [-1, 1]$, the following bound applies:*

$$f(x; \alpha, \beta, 1, 0; S0) \leq f(x; \alpha, \beta, 1, 0; S0)|_{x=m(\alpha,0;0)} = \frac{\Gamma(1 + 1/\alpha)}{\pi}. \tag{4.223}$$

Fofack and Nolan (1999) present numerical analysis of the behaviour of the mode as a function of tail index α and skewness parameter β. We also note that Hall (1984) was able to show that in the case of the quasi-symmetric α-stable severity model with β in the neighbourhood of the origin, ($\beta \neq 0$), one can derive asymptotic formulae for the unique mode of a stable law when the law is close to being symmetric.

In the following section, we will develop properties of the tail index and we will come back to relate these results to properties of the tail estimator and Hill estimators developed in the earlier discussion on extreme value theory (EVT).

4.10 Asymptotic Tail Properties of α-Stable Severity Models and Rates of Convergence to Paretian Laws

There are many results known about the tail behaviour of the α-stable model, as the application of such models in practice typically starts with discussion on the potential for heavy tails that such models provide. We start the discussion of the asymptotic tail properties of the α-stable family of models by observing the following feature of the tail function of a stable distribution under Nolan's $S1$ parameterization

$$\begin{aligned} \lim_{\lambda \to \infty} \lambda^\alpha \mathbb{P}\mathrm{r}[X > \lambda] &= C_\alpha \frac{(1+\beta)}{2} \gamma^\alpha, \\ \lim_{\lambda \to \infty} \lambda^\alpha \mathbb{P}\mathrm{r}[X < -\lambda] &= C_\alpha \frac{(1-\beta)}{2} \gamma^\alpha, \end{aligned} \tag{4.224}$$

where $C_\alpha = \left(\int_0^\infty x^{-\alpha} \sin x \; dx\right)^{-1}$.

This result illustrates one of the key features of α-stable models in applications in OpRisk, namely, that they admit heavy tails that are asymptotically Paretian. This property can be shown by considering the asymptotic series expansions from Bergstrom (1953) which as discussed in McCulloch (1997) and implies that as $x \to \infty$, then the tails of the α-stable model behave as given in Lemma 4.14. This shows that when the characteristic exponent is less than 2, the tails of α-stable severity models behave as a power law in $x^{-\alpha}$.

As discussed in Feller (1966, p. 175) and Fofack and Nolan (1999) in the special case of the Gaussian distribution $\alpha = 2$ and $\beta = 0$, the tail behaviour of the severity distribution, $X \sim Normal(0, 1)$, is approximated asymptotically $(x \to \infty)$ by the expression

$$\mathbb{P}\mathrm{r}[X > x] \sim \frac{\exp(-x^2/2)}{x\sqrt{2\pi}}. \tag{4.225}$$

Extending this result, it was shown by Lévy (1925) that the tails of non-Gaussian stable distributions are asymptotically equivalent to the Pareto law. Furthermore, one can present the following results on the tail behaviour of the univariate α-stable distribution for an analytic representation of the tail probability of the stable distribution in Lemma 4.14.

Lemma 4.14 *Given a random variable $X \sim S(\alpha, \beta, \gamma, \delta; S0)$, then as $x \to \infty$, one can write the limiting tail distribution*

$$\begin{aligned} \mathbb{P}r[X > x] &\sim \gamma^{\alpha} C_{\alpha}(1+\beta)x^{-\alpha}, \text{ asx} \to \infty, \\ f_X(x|\alpha, \beta, \gamma, \delta; S0) &\sim \alpha\gamma^{\alpha} C_{\alpha}(1+\beta)x^{-(\alpha+1)}, \text{ asx} \to \infty, \end{aligned} \tag{4.226}$$

where $C_{\alpha} = \sin(\pi\alpha/2)\,\Gamma(\alpha)/\pi$. This result follows from Nolan (2013, theorem 1.12).

In addition, it was shown that one could obtain the following tail behaviours for the B-type parameterizations of Zolatarev for the density and distribution, see Zolotarev (1986, theorems 2.5.4 and 2.5.5).

Proposition 4.10 (Zolotarev's Asymptotic Series Expansions Type-B) *Consider the standardized B-type α-stable random variable, standardized such that $\gamma = 1$ and $\delta = 0$. The density can be evaluated asymptotically with $x \to \infty$ by*

$$f_X(x; \alpha, \beta; B) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{k!} P_k(\ln x) x^{-k-1} \tag{4.227}$$

for $\alpha = 1$ and $\beta \in (-1, 1]$ and with

$$\begin{aligned} P_k(x) &= \sum_{l=0^k} r_{lk} x^l, \\ r_{lk} &= \sum_{m=l}^{k} C_m^k \, C_l^m (-1)^{m-l} \Gamma^{(m-l)}(1+n) \beta^m \left(\frac{\pi}{2}(1+\beta)\right)^{k-m} \sin\left(\frac{\pi}{2}(k-m)\right), \end{aligned}$$

where $\Gamma^{(m)}(x)$ is the mth derivative of the Gamma function. The distribution can be evaluated asymptotically with $x \to \infty$ by

$$F_X(x; \alpha, \beta; B) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{k!} P_k^*(\ln x) x^{-k} \tag{4.228}$$

for $\alpha = 1$ and $\beta \in (-1, 1]$ and with $P_k^*(x)$ evaluated with coefficients r_k^* given by

$$P_k^*(x) = \sum_{l=0^k} r_{lk}^* x^l,$$

$$r_{lk}^* = \sum_{m=l}^{k} C_m^k \, C_l^m (-1)^{m-l} \Gamma^{(m-l)}(n) \beta^m \left(\frac{\pi}{2}(1+\beta)\right)^{k-m} \sin\left(\frac{\pi}{2}(k-m)\right).$$

In addition, the tail of the distribution function can be evaluated asymptotically according to the following series expansion as $x \to \infty$,

$$1 - F_X(x; \alpha, \beta; B) \sim \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{\Gamma(\alpha k)}{\Gamma(k+1)} \sin \frac{\pi}{2} k(2-\alpha)(1+\beta) x^{-k\alpha} \tag{4.229}$$

for $\alpha > 1$ and $\beta \neq 1$.

In Example 4.10, we utilize this definition of the tail properties of the α-stable distribution and density functions to characterize an α-stable distribution with respect to the tail exponent in terms of a function of regular variation.

EXAMPLE 4.10

Here we show that the density function of a univariate α-stable random variable X is a member of the class of regularly varying functions with index of variation given by the tail exponent α. To demonstrate this, we consider the survival function that results asymptotically and demonstrate that it is a member of the class of regularly varying functions with $\overline{F} \in RV_\alpha$. This holds because for all values $b > 0$, there exists a real constant α for which the following limit is satisfied:

$$\begin{aligned} \lim_{x \to \infty} \frac{\overline{F}_X(bx)}{\overline{F}_X(x)} &= \lim_{x \to \infty} \frac{\gamma^\alpha c_\alpha (1+\beta)(bx)^{-\alpha}}{\gamma^\alpha c_\alpha (1+\beta) x^{-\alpha}} \\ &= \lim_{x \to \infty} \frac{(bx)^{-\alpha}}{x^{-\alpha}} = b^{-\alpha}. \end{aligned} \tag{4.230}$$

■

This has many implications for the properties of these random variables as well as the simulation and estimation that will be discussed. In particular, the feature that for $\alpha < 2$ results in the tails of α-stable severity models behave as a power law in $x^{-\alpha}$ means that α-stable models will have infinite absolute population moments of order greater than or equal to α, as given in Lemma 4.15. Before presenting this result for the fractional lower order moments (FLOMs), we note that one may write the fractional moment of a loss random variable $X \sim S(\alpha, \beta, \gamma, \delta; B)$ according to the following integral

$$\mathbb{E}[|X|^s] = 2\Gamma(1+s) \sin\left(\frac{2\pi}{s}\right) \int_0^\infty (1 - \mathscr{R}\mathrm{e}[\Phi_X(\theta)]) \theta^{s-1} \, d\theta. \tag{4.231}$$

This led Zolatarev to derive the following bound on the 'centralized' absolute moments given in Zolotarev (1986, Property 2.5), which shows that given a stable loss random variable

$X \sim S(\alpha, \beta, \gamma, \delta; B)$, for any $s \in [0, \alpha)$, one has

$$\mathbb{E}[|X - \text{Median}(X)|^s] \leq 4(2\gamma)^{s/\alpha}\Gamma\left(1 - s/\alpha\right)\Gamma(s)\sin\left(\frac{\pi}{2}s\right). \tag{4.232}$$

The consequence of this result is that $\mathbb{E}[|X|^s] < \infty$ for all $s \in [0, \alpha)$. One can further obtain the following result known as the fractional lower order moment (FLOM).

Lemma 4.15 *Consider the stable distributed loss random variable $X \sim S(\alpha, 0, \gamma, \delta)$, the FLOMs are given for $\alpha \in (0, 2)$ by*

$$\mathbb{E}[|Y|^p] = \begin{cases} \infty, & p \geq \alpha, \\ C(p, \alpha)\gamma^{p/\alpha}, & 0 < p < \alpha, \end{cases} \tag{4.233}$$

with

$$C(p, \alpha) = \frac{2^{p+1}\Gamma\left((p+1)/2\right)\Gamma\left(-p/\alpha\right)}{\alpha\sqrt{\pi}\Gamma\left(-p/2\right)}.$$

In general, it is precisely these tail properties of the α-stable model that make it of interest in OpRisk modeling for severity models for high consequence yet rare events. However, as discussed in Fofack and Nolan (1999) from the practical perspective, it is extremely important to understand the features of the α-stable model in order to appreciate the rate at which these tail approximations become reliable as a function of the tail index α and the skewness parameter β. Hall (1984) and Fofack and Nolan (1999) consider the rates of convergence of the α-stable law, with particular focus on the properties of these asymptotic expansions and when they are practically relevant to be applied. This will be of direct importance when such asymptotics are developed for estimation of risk measures in OpRisk. This is equivalent to addressing the question, at what point in the support of the α-stable distributed random variable does the Paretian tale approximation become accurate, as a function of α and β.

It was found that the general behaviour of the accuracy of the tail approximation as a function of the tail index and skewness, even in the restricted case of symmetric α-stable models is very complex to characterize, see discussions in Fofack and Nolan (1999).

Remark 4.29 *In general, it has been observed numerically by Fofack and Nolan (1999) that the distribution functions of the α-stable laws will approach the Paretian limit faster than the density functions will as $x \to \infty$. Consequently, inference and estimation of tail index and risk measures under such models will be more accurate when working with distribution functions.*

Remark 4.30 *Importantly, for OpRisk applications, it was observed by Fofack and Nolan (1999) that when there is a significant skew present, such as $|\beta| \to 1$, then the lighter tail takes a very long time to reach the Paretian limit.*

Remark 4.31 *As the tail index $\alpha \uparrow 2$, the Paretian tail behaviour does not occur until the tail quantile becomes extremely large. This makes sense intuitively as this is approaching the Gaussian type tail behaviour.*

It is discussed in numerous places in the literature that the consequences of this complex behaviour of the tails of the α-stable distribution as a function of α and β has significant implications for the practitioner. In particular, the procedures described in the EVT chapter

related to estimation of the tail index via a Hill estimator Hill (1975) which can be unreliable in certain members of the family of sub-exponential models, such as the α-stable sub-family. This has been discussed at length in Fofack and Nolan (1999), Resnick (1997) and McCulloch (1997).

We conclude this section with basic discussion about aspects of what is known theoretically about rates of convergence of α-stable laws. Some of the early work on this was developed by Satybaldina (1972), Butzer and Hahn (1978), Hall (1981a, 1981b, 1982, 1984), Egorov (1980) and more recently by Kuske and Keller (2000) amongst others.

It is known that when one considers convergence of a Normal distributional law, such that observed losses in the risk process being aggregated to create S_n given in Equation 2.5 satisfy conditions required, for example, in the Lindenberg CLT, then the uniform rate of convergence is completely determined by the rate of convergence at points near the origin up to terms of order $n^{-1/2}$. However, as is clear from the previous section and discussions on α-stable models and their properties, the rates of convergence of α-stable laws is far from similar to Normal law behaviour.

In Hall (1984), results for the analogue of the Normal convergence in an α-stable context were studied, that is, the symmetric setting in which $\beta = 0$ was the focus and results on rates of convergence were generalized from the Normal setting to the α-stable setting.

We consider the sequence $X_1, X_2, \ldots$ of independent identically distributed random variables with distribution F in the α-stable domain of attraction, $F \in \mathscr{D}(G) \subseteq \mho$. Furthermore, assume that each random variable is scaled and located in such a manner that the normed sum

$$S_n = n^{-1/\alpha} \sum_{i=1}^{n} X_i \tag{4.234}$$

has a limit distribution G with characteristic function $\exp(-|\theta|^\alpha)$. Denote c as the positive constant defined according to $1 - G(x) \sim cx^{-\alpha}$ as $x \to \infty$. It was shown in Ibragimov *et al.* (1971) that one has in this scenario that $x^\alpha F(-x) \to -c$ and $x^\alpha(1 - F(x)) \to c$ as $x \to \infty$.

In this case, it was proven in Hall (1984) that, under assumptions on the regularity of the tail sum and the tail difference given in Definition 4.29, the rate of convergence given in Theorem 4.25 holds.

Definition 4.29 *The tail sum is defined to be*

$$S(x) = 1 - F(x) + F(-x) - 2cx^{-\alpha}, \tag{4.235}$$

and the tail difference is given by

$$D(x) = 1 - F(x) - F(-x). \tag{4.236}$$

The premise of the α-stable rate of convergence result presented in Hall (1984), for convergence of sequences of suitably standardized linear combinations of random variables, whose distribution is in the domain of normal attraction of a α-stable law, to a limiting distribution G, is based upon imposing conditions on the regularity of the tail behaviour of the distribution F. Then under these assumptions, which in most cases are satisfied if one requires that the distribution F is regularly varying, one attains the ability to study the rate of convergence at particular points whilst extrapolating from these pointwise convergence studies, results about rates of convergence uniformly on the whole real line.

In effect, they show that when one considers certain assumptions on the regular variation of the tails of the severity distribution that is generating the loss process, then it is possible to study the rates of convergence at specific isolated points, such as around the mode, and these convergence behaviours may be extrapolated to inform rates uniformly on the support of the distribution.

One approach to understanding this is to consider the condition on the tail sum and tail difference, given in Equation 4.237.

$$\int_1^{\infty} x^{\alpha(1+d)-1}\{|S(x)| + |D(x)|\}dx < \infty. \tag{4.237}$$

In many scenarios, this acts as a sufficient condition to state

$$\sum_{n=1}^{\infty} n^{-1+d} \sup_{x\in\mathbb{R}} |\mathbb{P}\mathrm{r}[S_n \leq x] - G(x)| < \infty, \tag{4.238}$$

which provides the series measure of the convergence rate of the supremum, over all $x \in \mathbb{R}$, absolute distance between the sequence of normalized sums S_n to converge to the limiting distribution $G(x)$. This is under the constraints that $0 < d < -1 + 2/\alpha$ for $\alpha \in [1, 2)$ or $0 < d < \alpha^{-1} \min(1, 2\alpha) - 2$ for $\alpha \in (0, 1)$.

In effect, the result provided by Hall (1984) generalize this convergence from the supremum over the support of the distribution $G(x)$ to a result on the maximum over two points, whilst also showing the necessity of Equation 4.238 in this setting to satisfying the condition in Equation 4.237.

Theorem 4.25 *Assume that the tail sum and tail difference are ultimately monotone. Furthermore, assume the following regularity properties apply,*

$$\begin{cases} \int_1^{\infty} \left(|S(x)| + |D(x)|\right) dx < \infty, & \text{if } \alpha = 1, \\ \lim\limits_{z\to 0} \dfrac{\left(\int_z^{\infty} |S(x)|dx\right)^2}{z^{\alpha}\int_1^z x|S(x)|dx + z^3\int_z^{\infty} x^{-\alpha-2}|S(x)|dx} \to 0, & \text{if } 1 \leq \alpha < 2, \\ \lim\limits_{z\to\infty} \dfrac{\int_z^{\infty} x^{-1}|S(x)|dx}{z^{-2}\int_1^z x|S(x)|dx + z^{\alpha+1}\int_z^{\infty} x^{-\alpha-2}|S(x)|dx} < \infty, & \text{if } 0 < \alpha < 1, \\ \lim\limits_{z\to\infty} \dfrac{\int_z^{\infty} x^{-1}|D(x)|dx}{z^{-1}\int_1^z |D(x)|dx + z^{\alpha}\int_z^{\infty} x^{-\alpha-1}|D(x)|dx} < \infty, & \text{if } 0 < \alpha < 1, \end{cases}$$

then the following rate of big-oh convergence is known,

$$n\{|S(n^{1/\alpha})| + |D(n^{1/\alpha})|\} = \mathcal{O}\{\Delta_n(S) + n^{1-\kappa/\alpha}\}, \tag{4.239}$$

where $\Delta_n(S) = \max\limits_{x\in S} |\mathbb{P}\mathrm{r}[S_n \leq x] - G(x)|$ *with* $S = \{x_1, x_2\}$ *in which one of the two conditions holds: (i)* $x_1 = 0$ *and* x_2 *is non-zero or (ii)* $x_1 < 0 < x_2$ *and these points belong to an open interval including the origin,* $\{x_1, x_2\} \in S_x$*, in which strict unimodality is satisfied for the density of the limiting distribution* $g(x)$*.*

Remark 4.32 *These assumed conditions are constraints only on the regularity of the tail sum and difference, except in the special case of* $\alpha = 1$*, and not on their rate of convergence to zero. Furthermore, if* $\alpha \neq 1$ *and* $|S(x)|$ *and* $|D(x)|$ *are regularly varying, then the last three conditions are satisfied.*

CHAPTER FIVE

Flexible Heavy-Tailed Severity Models: Tempered Stable and Quantile Transforms

This chapter develops further the flexible α-stable severity models discussed previously by extending them into the classes of tempered stable (TS) models and quantile transform models. In particular, we present advanced discussion for risk modellers relating to flexible distributions that complements the flexible severity distributions typically utilized in OpRisk, such as those discussed in Cruz *et al.* (2015, chapter 9). In particular, this chapter will provide details on how to model losses in OpRisk settings with heavy-tailed severity models such as the following families:

- the TS family with exponential tilting and its relationship to the generalized inverse Gaussian (GIG) and generalized hyperbolic families discussed in Cruz *et al.* (2015, chapter 9);
- the generalized tempered stable (GTS) family and the modified tempered stable (MTS) family;
- mixture variants that generalize those presented in the generalized hyperbolic family such as the normal-modified stable family;
- quantile function flexible severity models based on Tukey's transform class.

5.1 Tempered and Generalized Tempered Stable Severity Models

In the following sections, we discuss the consideration of modifications to the α-stable family to offer additional flexibility with regard to the modeling of skewness and kurtosis in a dataset.

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

Primarily, the approaches considered will be based on taking the Lévy–Khintchine representation of the characteristic function, for an infinitely divisible loss random variable, given in Definition 4.16. Then given this representation for the severity model considered in the OpRisk loss distribution approach (LDA) structure one may proceed under two frameworks, the TS models or the related class of models that are termed the tempered infinitely divisible (TID) severity models, see discussions on the relationships in Bianchi *et al.* (2011). In particular, a key property that can distinguish between the TS and TID models is that whilst the TS models are constructed from α-stable models, they admit all moments and some limited number of exponential moments, whilst the TID family can admit all exponential moments if suitably designed. In fact, one should take particular care in the understanding because there are many variations of TS and TID members, some of which are jointly labelled as TS but actually turn out to be TID sub-families. Generally, the literature is moving towards denoting the TS family as those of the Rosiński form, after the seminal work of Rosiński (2007) and those that are not represented in this form may be considered as candidates from the TID family, examples include the MTS and normal modified tempered stable (NMTS) families of Barndorff-Nielsen & Shephard (2001). In the following discussion, we explain the numerous representations of the TS and TID members that have been proposed and can be considered as useful and flexible severity models for OpRisk LDA models.

Starting from the premise that one simply *tempers* the spectral measure, then new classes of severity models emerge from for instance the α-stable family of models. The development of such a family of models first arose from the work of Tweedie (1984) and Hougaard (1986). More recently, these classes of models have found a revival in the financial economics and actuarial literatures. These models have generated significant interest in recent years, and they have been developed under the following model names:

- Truncated Lévy Flights, see Novikov (1994), Koponen (1995), Mantegna & Stanley (1994) and Chan (1999);
- KoBoL distributions, see Boyarchenko & Levendorskii (2000) and Matacz (2000);
- Variance gamma processes, see Madan *et al.* (1998);
- CGMY processes, see Poirot & Tankov (2006), Carr *et al.* (2002) and Cont *et al.* (1997);
- TS processes, see Rosiński (2007) and Kienitz (2010), which have at least two variants, the classical tempered stable (CTS) and the rapidly decreasing tempered stable (RDTS) models see discussion in Shin Kim *et al.* (2010);
- Normal modified and TS processes, see Barndorff-Nielsen & Shephard (2001);
- TID models, see discussion in Bianchi *et al.* (2011).

5.1.1 UNDERSTANDING THE CONCEPT OF TEMPERING STABLE SEVERITY MODELS

We begin by defining what tempering of the spectral measure corresponds to, as presented in Proposition 5.1. Unlike the stable distribution, the TS severity model can have all moments finite including exponential moments of some order. The idea of tempering a stable distribution can be related back to the concept of tilting of a distribution. Throughout this chapter, the notion of tempering will be discussed so it seems natural to open the discussion with the definition of both a discrete Esscher transform and a continuous Esscher transform given in Definition 5.1, also commonly known in statistics as an 'exponential tilting' and used widely in

developing asymptotic series expansions for distributions and densities such as the Edgeworth or saddle point types, see further details in Small (2010) and references therein.

Definition 5.1 (Esscher Transform or Exponential Tilting) *Consider a continuous random variable X defined with respect to a probability measure $\mathbb{P}$ and a non-zero constant real number h such that $\mathbb{E}_{\mathbb{P}}[\exp(hX)]$ exists. One can then define the Esscher transform, denoted by $\mathscr{E}_h[\cdot]$, of the original probability measure $\mathbb{P}$ for X in terms of an equivalent new probability measure $\mathbb{Q} \sim \mathbb{P}$ (same null sets) with the following properties:*

1. $\mathscr{E}_{h_1}\mathscr{E}_{h_2}[\mathbb{P}] = \mathscr{E}_{h_1+h_2}[\mathbb{P}]$,
2. $\mathscr{E}_h^{-1}[\mathbb{P}] = \mathscr{E}_{-h}[\mathbb{P}]$.

If the measure $\mathbb{P}$ that characterizes random variable X admits a Radon–Nikodym derivative with respect to a suitable measure ν given by a density $f_X(x)$, then we see that the Esscher transform of the density, denoted by $\mathscr{E}_h[f_X(x)]$, will be a new density given by

$$f(x;h) := \mathscr{E}_h[f_X(x)] = \frac{\exp(hx)f_X(x)}{\int_{-\infty}^{\infty}\exp(hx)f_X(x)\,dx}. \tag{5.1}$$

In Rosiński (2007), a very nice explanation of this was developed, which we discuss briefly later. In explaining how tempering works, one can start from the well-known idea of *tilting* of a density function and can show how this can be reworked into a tilting in the spectral measure. Consider a severity density $f(x)$ with support on $\mathbb{R}^+$, which we assume for the sake of this illustration, can be represented in an exponential family form with $f(x) = \exp(k(x))$ for some suitable function $k(x)$ depending on the parameterization and family member of the exponential family. Furthermore, let us assume that the random variable X with density $f(x)$ is infinitely divisible. Then define the *tilted* density, with respect to a tilting parameter $\theta > 0$, by the new density

$$f(x;\lambda) = \frac{1}{Z(\lambda)}\exp(-\lambda x)f(x) = \exp(-\lambda x + z(\lambda) + k(x)), \tag{5.2}$$

where the normalizing constant $Z(\lambda)$ is the Laplace transform of the density $f(x)$ and $Z(\lambda) = \exp(-z(\lambda))$. In this setting, one sees that the tilted distribution forms a one-parameter exponential family member.

Now, taking the Laplace transform of the density (to obtain the moment-generating function (MGF)), one obtains

$$\mathscr{F}(\theta;\lambda) = \exp(-l(\theta+\lambda) - l(\lambda)). \tag{5.3}$$

Given the infinite divisibility of X, one can, therefore, represent the component of the Laplace transform $l(\theta)$ by the following generic form

$$l(\theta) = \theta b + \int_0^{\infty}[1 - \exp(-\theta x)]W(dx). \tag{5.4}$$

Note: this uses the Lévy–Khintchine representation of an infinitely divisible distribution for the MGF and here $W(dx)$ represents the Lévy measure on $\mathbb{R}^+$ with $b \geq 0$.

Therefore, substituting this representation into the MGF expression, one obtains

$$\mathscr{F}(\theta;\lambda) = \exp\left[-\theta b - \int_0^\infty (\exp(-\theta x) - 1)\exp(-\lambda x) W(dx)\right]. \tag{5.5}$$

Now it should be apparent that the tilting that is applied directly to the distribution function can be re-expressed directly in terms of a tempering of the Lévy measure. Hence, the name tempering of the Lévy measure, which corresponds to a tilting of the density. It is also clear, therefore, that this adjustment to the density is analogous to the Esscher transform discussed in incomplete market pricing of contingent claims, see Cruz *et al.* (2015, Insurance Chapter).

Hence, in general, one may consider tempering of a Lévy measure of an infinitely divisible distribution as detailed in Proposition 5.1.

Proposition 5.1 (Tempering Lévy Measure of an Infinite Divisible Severity Model) *Given a distribution function F, which is infinitely divisible, then the logarithm of its characteristic function admits a Lévy–Khintchine representation given generally by*

$$\Phi_X(\theta) = \exp(\psi(\theta)), \quad \theta \in \mathbb{R}, \tag{5.6}$$

where the exponent of the characteristic function $\psi(\theta)$ is given by

$$\psi(\theta) = \mathrm{i}a_g\theta - \tfrac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{\mathrm{i}\theta x} - 1 - \mathrm{i}\theta g(x))\, W(dx), \tag{5.7}$$

for some truncation function $g : \mathbb{R} \to \mathbb{R}$, which is measurable and satisfies for all θ the condition

$$\int |\exp(\mathrm{i}\theta x) - 1 - \mathrm{i}\theta g(x)| W(dx) < \infty, \tag{5.8}$$

with $a_g \in \mathbb{R}$ depending on the choice of truncation function and $\sigma \geq 0$ and the Lévy measure W which satisfies

$$\int_{\mathbb{R}\setminus\{0\}} \min\{x^2, 1\} W(dx) < \infty. \tag{5.9}$$

If one considers a tempering parameter $\lambda \in \mathbb{R}$, then it is possible to show that using λ, one obtains finite exponential moments with

$$\mathbb{E}[\exp(\lambda X)] < \infty, \tag{5.10}$$

if and only if the following condition is satisfied for the tempered spectral measure

$$\int_{|x|\geq 1} \exp(\lambda x) W(dx) < \infty, \tag{5.11}$$

and under this tempering, one obtains

$$\mathbb{E}[\exp(\lambda X)] = \exp(\psi(-\mathrm{i}\lambda)). \tag{5.12}$$

5.1.2 FAMILIES AND REPRESENTATIONS OF TEMPERING IN STABLE SEVERITY MODELS

Using this notion of tempering, one may then define the different classes of TS severity models for consideration in OpRisk modeling. Again, in the literature, on TS distributions, as was the case in the stable model literature, there are variations on what is defined to be the TS representation. In introducing the TS distribution, we find it fitting to first discuss some of the results derived in the detailed analysis provided in Rosiński (2007). In particular, this paper derives the following representation of the Lévy measure of a TS distribution in terms of the Rosiński spectral measure, see details of derivations in Rajput & Rosiński (1989) and Sato (1999). Note that the results later are expressed for univariate settings of relevance to this chapter, although there are natural extensions to multivariate space developed.

Theorem 5.1 (Lévy and Rosiński Measures for Tempered Stable Severity Model) *The Lévy measure $W(dx)$ of a TS distribution can be expressed in the following form for some Borel set $A \in \mathscr{B}(\mathbb{R})$*

$$W(A) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathbb{I}_A(\theta x)\, \theta^{-\alpha-1} \exp(-\theta) d\theta R(dx), \tag{5.13}$$

where $R(dx)$ is a unique measure (Rosiński spectral measure) on $\mathbb{R}$ such that it satisfies

$$R(\{0\}) = 0, \quad \textit{and}$$
$$\int_{-\infty}^{\infty} \min(x^2, x^\alpha) R(dx) < \infty, \tag{5.14}$$

where a proper TS distribution will satisfy

$$\int_{-\infty}^{\infty} x^\alpha R(dx) < \infty. \tag{5.15}$$

The parameter α in this representation is the direct analogue of the α tail index parameter in the stable severity models presented previously in Chapter 4.

Remark 5.1 *The unique measure $R(dx)$ is known as the spectral measure of the TS distribution.*

Another way of thinking about the Rosiński measure is to consider writing the Lévy measure $W(dx)$ of the TS severity model in polar co-ordinates as shown in Definition 5.2, see discussion in Bianchi *et al.* (2010).

Definition 5.2 (Lévy Measure to Rosiński Measure of TS Models) *Consider the loss random variable $X \in \mathbb{R}^d$ (e.g. with $d = 1$) and having a TS distribution with tail index $\alpha \in (0, 2)$, which is infinitely divisible and with Lévy measure $W(dx)$. Then the Lévy measure $W(dx)$ can be written in polar co-ordinates via*

$$W(dr, du) = r^{-\alpha-1} q(r, u) dr \sigma(du), \tag{5.16}$$

for some finite measure $\sigma \in \mathbb{S}^{d-1}$ and tempering function

$$q : (0, \infty) \times \mathbb{S}^{d-1} \mapsto (0, \infty), \tag{5.17}$$

such that the Borel function $q(\cdot, u)$ is completely monotone and $q(\infty, u) = 0$ for all $u \in \mathbb{S}^{d-1}$.

Notice that in this family of TS models, the decay of the exponential arguments in the Lévy measure are linear; this results in a slower tail decay; it was realized that one may develop a sub-family of TS models with shorter tails by increasing the exponential decay arguments to be quadratic monomials. Hence, one can then define the family of characteristic functions for the TS distribution according to the following representation given in Definition 5.3, see derivations in Rosiński (2007).

Definition 5.3 (Representation 1 of Tempered Stable Severity Models) *An infinitely divisible loss random variable X has a TS distribution $X \sim TS(\alpha, R(dx), b; R1)$ if its characteristic function given by*

$$\Phi_X(\theta) = \exp(\psi(\theta)), \tag{5.18}$$

where the exponent function is given by one of the following two forms, with Lévy measure as defined in Theorem 5.1.

- *If one the following three cases is satisfied:*
  1. $\alpha \in (1, 2)$*;*
  2. $\alpha = 1$ *and*
  $$\int_{|x|>1} x \ln(x) R(dx) < \infty; \tag{5.19}$$
  3. $\alpha \in (0, 1)$ *and*
  $$\int_{|x|\leq 1} xR(dx) < \infty, \tag{5.20}$$

  then
  $$\psi(\theta) = \begin{cases} \mathrm{i}\theta b + \int_{-\infty}^{\infty} \Gamma(-\alpha) \left[(1 - \mathrm{i}\theta x)^{\alpha} - 1 + \mathrm{i}\alpha\theta x\right] R(dx), & \alpha \neq 1, \\ \mathrm{i}\theta b + \int_{-\infty}^{\infty} [(1 - \mathrm{i}\theta x) \ln(1 - \mathrm{i}\theta x) + \mathrm{i}\theta x] R(dx), & \alpha = 1, \end{cases} \tag{5.21}$$

  with $b = \int_{-\infty}^{\infty} xF(dx)$.
- *If* $\alpha \in (0, 1)$ *and*
$$\int_{|x|\leq 1} |x| R(dx) < \infty, \tag{5.22}$$

  then
$$\psi(\theta) = \mathrm{i}\theta b_0 + \int_{-\infty}^{\infty} \Gamma(-\alpha)[(1 - \mathrm{i}\theta x)^{\alpha} - 1] R(dx), \tag{5.23}$$

  with $b_0 \in \mathbb{R}$.

From this general representation of the TS family of distributions we note several other representations of this family of severity models in terms of their characteristic function and Lévy triplet. We begin with the representation adopted in Shin Kim *et al.* (2010) where the CTS process is defined by the Lévy triplet given in Definition 5.4.

Definition 5.4 (Representation 2 of Tempered Stable Severity Model Lévy Triplet (CTS)) *An infinitely divisible loss random variable X has a CTS distribution $X \sim CTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R2)$ if its characteristic function is given by*

$$\Phi_X(\theta) = \exp(\psi(\theta)) \tag{5.24}$$

with

$$\begin{aligned}\psi(\theta) &= \mathrm{i}\theta m - \mathrm{i}\theta\Gamma(1-\alpha)[c_+\lambda_+^{\alpha-1} - c_-\lambda_-^{\alpha-1}] \\ &\quad + c_+\Gamma(-\alpha)[(\lambda_+ - \mathrm{i}\theta)^\alpha - \lambda_+^\alpha] + c_-\Gamma(-\alpha)[(\lambda_- + \mathrm{i}\theta)^\alpha - \lambda_-^\alpha],\end{aligned} \tag{5.25}$$

and its Lévy triplet is therefore given by (a_g, σ^2, W) *with* $\sigma^2 = 0$ *and*

$$W(dx) = [c_+ \exp(-\lambda_+ x)\mathbb{I}_{x>0} + c_- \exp(-\lambda_-|x|)\mathbb{I}_{x<0}]\frac{dx}{|x|^{\alpha+1}}, \tag{5.26}$$

and

$$a_g = m - \int_{|x|>1} xW(dx). \tag{5.27}$$

In addition, Shin Kim *et al.* (2010) also define the RDTS process by the Lévy triplet given in Definition 5.5.

Definition 5.5 (Representation 3 of Tempered Stable Severity Model Lévy Triplet (RDTS)) *An infinitely divisible loss random variable* X *has an RDTS distribution* $X \sim RDTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R3)$ *if its Lévy triplet is given by*

$$\begin{aligned}\sigma &= 0, \\ W(dx) &= \left[c_+ \exp\left(-\frac{1}{2}\lambda_+^2 x^2\right)\mathbb{I}_{x>0} + c_- \exp\left(-\frac{1}{2}\lambda_-^2|x|^2\right)\mathbb{I}_{x<0}\right]\frac{dx}{|x|^{\alpha+1}}, \\ a_g &= m - \int_{|x|>1} xW(dx).\end{aligned}$$

The characteristic function of the RDTS family of severity models is then given by

$$\Phi_X(\theta) = \exp[\mathrm{i}\theta m + c_+ G(\theta; \alpha, \lambda_+) + c_- G(\theta; \alpha, \lambda_-)], \tag{5.28}$$

with the function

$$\begin{aligned}G(\theta; \alpha, \lambda) &= 2^{-\frac{\alpha}{2}-1}\lambda^\alpha\Gamma\left(-\frac{\alpha}{2}\right)\left[\mathrm{M}\left(-\frac{\alpha}{2}, \frac{1}{2}; \frac{x^2}{2\lambda^2}\right) - 1\right] \\ &\quad + 2^{-\frac{\alpha}{2}-\frac{1}{2}}\lambda^{\alpha-1}x\Gamma\left(\frac{1-\alpha}{2}\right)\left[\mathrm{M}\left(\frac{1-\alpha}{2}, \frac{3}{2}; \frac{x^2}{2\lambda^2}\right) - 1\right],\end{aligned} \tag{5.29}$$

where M is the confluent hypergeometric function, see Andrews (1985) with Slater (1960) and the representation can take many forms for instance

$$\mathrm{M}(a, b; x) = \sum_{n=0}^{\infty} \frac{a^{(n)}x^n}{b^{(n)}n!}, \tag{5.30}$$

where $a^{(n)}$ *is the rising factorial given by*

$$\begin{aligned}a^{(0)} &= 1, \\ a^{(n)} &= a(a+1)(a+2)\ldots(a+n-1).\end{aligned} \tag{5.31}$$

In Bianchi *et al.* (2011), the RDTS family of models is shown to be a member of the TID family with a Rosiński spectral measure corresponding to the Lévy measure in Definition 5.5 comprised the sum of two Dirac measures that are weighted appropriately. The following Rosiński spectral measure for the RDTS family is given in Lemma 5.1, see Bianchi *et al.* (2011, Example 1).

Lemma 5.1 *Consider a loss random variable X with RDTS distribution $X \sim RDTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R3)$ with a Lévy spectral measure given by*

$$\begin{aligned} \sigma &= 0, \\ W(dx) &= \left[c_+ \exp\left(-\frac{1}{2}\lambda_+^2 x^2 \right) \mathbb{I}_{x>0} + c_- \exp\left(-\frac{1}{2}\lambda_-^2 |x|^2 \right) \mathbb{I}_{x<0} \right] \frac{dx}{|x|^{\alpha+1}}, \\ a_g &= m - \int_{|x|>1} x W(dx). \end{aligned} \tag{5.32}$$

Then the corresponding Rosiński measure can be obtained by transforming to Polar co-ordinates to obtain the Rosiński measure

$$R(A) = c_+ \int_A \lambda_+^\alpha \delta_{\lambda_+^{-1}}(x)\, dx + c_- \int_A \lambda_-^\alpha \delta_{\lambda_-^{-1}}(x)\, dx, \tag{5.33}$$

where $\delta_{\lambda_+^{-1}}(x)$ and $\delta_{\lambda_-^{-1}}(x)$ are each Dirac delta functions taking value of unity if $x = \lambda_+^{-1}$ or $x = \lambda_-^{-1}$ respectively, otherwise they each take a value of zero.

Remark 5.2 *In the CTS and RDTS models, the parameters λ_+ and λ_- control the rate of decay of the positive and negative tails with*

- *the severity distribution is skewed to the left when $\lambda_+ > \lambda_-$;*
- *the severity distribution is skewed to the right when $\lambda_+ < \lambda_-$; and*
- *the severity distribution is symmetric when $\lambda_+ = \lambda_-$.*

In the CTS and RDTS families of severity models, the tail index of the distribution, denoted by α, is considered common on both the left and the right tails. This was generalized in the TS family in Definition 5.6 known as the 'generalized tempered stable' (GTS) family of severity models, see discussions in Poirot & Tankov (2006). Note, this family of GTS models is also known as the KoBoL family.

Definition 5.6 (Representation 4 of GTS Severity Model Spectral Measures) *A loss random variable X has a GTS representation $X \sim GTS(\alpha_+, \alpha_-, c_+, c_-, \lambda_+, \lambda_i; R4)$ is characterized by its Tempered Lévy spectral measure given by*

$$W(dx) = \frac{c_+ \exp(-\lambda_+ x)\, dx}{x^{1+\alpha_+}} \mathbb{I}_{x>0} + \frac{c_- \exp(-\lambda_- x)\, dx}{|x|^{1+\alpha_-}} \mathbb{I}_{x<0} \tag{5.34}$$

with parameters $c_- > 0$, $c_+ > 0$, $\lambda_- > 0$, $\lambda_+ > 0$, $\alpha_+ \in (0, 2)$ and $\alpha_- \in (0, 2)$.

Using this spectral measure representation of the class of TS severity models for consideration in OpRisk modeling, one then has the following representation of the Lévy–Khinitchine characteristic function, presented in Definition 5.7. Note that as discussed previously in the α-stable setting, we can again choose different choices for the truncation function.

Definition 5.7 (Lévy–Khintchine Characteristic Function for Tempered Stable Severity Models) *For a distribution function F which is infinitely divisible, the logarithm of its characteristic function admits a Lévy–Khintchine representation given generally by*

$$\Phi_X(\theta) = \exp(\psi(\theta)), \quad \theta \in \mathbb{R}, \tag{5.35}$$

where the exponent of the characteristic function $\psi(\theta)$ is given by

$$\psi(\theta) = \mathrm{i}a_g\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{\mathrm{i}\theta u} - 1 - \mathrm{i}\theta g(x))\, W(dx), \tag{5.36}$$

for some truncation function $g : \mathbb{R} \to \mathbb{R}$ which is measurable and satisfies for all θ the condition

$$\int |\exp(\mathrm{i}\theta u) - 1 - \mathrm{i}\theta g(x)| W(dx) < \infty, \tag{5.37}$$

with $a_g \in \mathbb{R}$ depending on the choice of truncation function and $\sigma \geq 0$ and the Lévy measure W which satisfies

$$\int_{\mathbb{R}\setminus\{0\}} \min\{x^2, 1\} W(dx) < \infty. \tag{5.38}$$

If the truncation function $g_0(x) = 0$ with $\alpha_+ = \alpha_- = \alpha \in (0, 1)$, then one obtains

$$\psi(\theta) = \mathrm{i}a_{g_0}\theta - \Gamma(-\alpha)[c_+((\lambda_+ - \mathrm{i}\theta)^\alpha - \lambda_+^\alpha) + c_-((\lambda_- + \mathrm{i}\theta)^\alpha - \lambda_-^\alpha)], \tag{5.39}$$

and if the truncation function is selected as $g_1(x) = x$ with $\alpha_\pm \neq 1$ and $\alpha_\pm \neq 0$, then one obtains

$$\psi(\theta) = \mathrm{i}a_{g_1}\theta - \Gamma(-\alpha_+)\lambda_+^{\alpha_+} c_+ \left[\left(1 - \frac{\mathrm{i}\theta}{\lambda_+}\right)^{\alpha_+} - 1 + \frac{\mathrm{i}\theta\alpha_+}{\lambda_+}\right]$$

$$+ \Gamma(-\alpha_-)\lambda_-^{\alpha_-} c_- \left[\left(1 + \frac{\mathrm{i}\theta}{\lambda_-}\right)^{\alpha_-} - 1 - \frac{\mathrm{i}\theta\alpha_-}{\lambda_-}\right]. \tag{5.40}$$

We also note the following relationship between the four representations presented and the model in the financial literature that goes by the name of the Carr, Geman, Madan and Yor (CGMY) model

Remark 5.3 *In the case that $c_+ = c_-$ and $\alpha_+ = \alpha_-$, one recovers the model in financial literature on options pricing known as the Carr, Geman, Madan and Yor (CGMY) model given by*

$$W(dx) = C\left[\frac{\exp(-Mx)\, dx}{x^{1+Y}}\mathbb{I}_{x>0} + \frac{\exp(-Gx)\, dx}{|x|^{1+Y}}\mathbb{I}_{x<0}\right], \tag{5.41}$$

where the variable C, G, M and Y have the following interpretations:

1. *C is a measure of overall levels of activity (in a diffusion context).*
2. *G is a measure of the skewness of the severity distribution.*
3. *M is a measure of skewness of the severity distribution.*
4. *Y is a measure of the fine (local) structure (in a diffusion context).*

In the GTS family, one can also recover the Kim-Rachevz (KR) distribution of Kim *et al.* (2008) which had a Rosiński measure and characteristic function representation developed in Proposition 5.2.

Proposition 5.2 (KR Sub-family of GTS Severity Models) *Consider the loss random variable X with a GTS severity model given by $X \sim GTS(\alpha_+, \alpha_-, c_+, c_-, \lambda_+, \lambda_i; R4)$. Considering the relabelled parameters $\alpha \in (0,2)$ with $\delta \in \mathbb{R}$, $k_+ > 0$, $k_- > 0$, $r_+ > 0$, $r_- > 0$ and $p_\pm \in (-\alpha, \infty) \notin \{-1, 0\}$, one can obtain the KR severity models as members of the GTS family if they have a Rosi'nski measure given by*

$$R(dx) = \left[k_+ r_+^{-p_+} |x|^{p_+ - 1} \mathbb{I}_{(0,r_+)}(x) + k_- r_-^{-p_-} |x|^{p_- - 1} \mathbb{I}_{(-r_-,0)}(x)\right] dx, \tag{5.42}$$

and a corresponding characteristic function given by

$$\Phi_X(\theta) = \exp(\psi(\theta)), \quad \theta \in \mathbb{R}, \tag{5.43}$$

with

$$\psi(\theta) = \frac{k_+ \Gamma(-\alpha)}{p_+} \left[2F_1\left(p_+, -\alpha; 1 + p_+; \mathrm{i} r_+ \theta\right) - 1 + \frac{\mathrm{i}\alpha p_+ r_+ \theta}{p_+ + 1}\right]$$

$$+ \frac{k_- \Gamma(-\alpha)}{p_-} \left[2F_1\left(p_-, -\alpha; 1 + p_-; \mathrm{i} r_- \theta\right) - 1 - \frac{\mathrm{i}\alpha p_- r_- \theta}{p_- + 1}\right] + \mathrm{i}\theta\delta, \tag{5.44}$$

with the mean of the KR distributions being given by δ.

In Jelonek (2012), a representation of a family of TS random variables is provided, which is a five-parameter representation of the TS characteristic function that was presented under a parameterization analogous to those derived from a Lévy–Khintchine representation to, for instance, the Nolan S1 parameterization. This provides the Representation 5 given in Definition 5.8 which looks like a five-parameter extension of the α-stable models discussed previously. Note that the representation adopted below has a slight change in parameter labels to associate the parameters with those of the Stable families presented previously. The additional parameter $\lambda > 0$ that controls the strength of the exponential tempering of the TS spectral measure $\exp(-\lambda|x|)$ has the following features: if $\lambda \approx 0$, then one recovers the α-Stable family of severity models; and large $\lambda >> 0$ produces a significantly lighter tailed severity model compared to the α-stable model with the same α tail index.

Definition 5.8 (Representation 5 of Tempered Stable Severity Model Lévy Triplet) *A loss random variable X has a TS distribution $X \sim TS(\alpha, \beta, \gamma, \delta, \lambda; R5)$ if characteristic function is given by*

$$\Phi_X(\theta) = \exp[\mathrm{i}\theta(\delta - g(\theta)) + \psi(\theta)] \tag{5.45}$$

with the functions

$$g(\theta) = \begin{cases} \alpha\left[\cos\left(\frac{\pi\alpha}{2}\right)\right]^{-1}\beta\gamma^{\alpha}\lambda^{\alpha-1}, & \alpha \neq 1, \\ -\frac{2}{\pi}\beta\gamma(\ln\lambda + 1), & \alpha = 1, \end{cases}$$

$$\psi(\theta) = \begin{cases} -\frac{1}{2}\left[\cos\left(\frac{\pi\alpha}{2}\right)\right]^{-1}\gamma^{\alpha}[(1+\beta)(\lambda - \mathrm{i}\theta)^{\alpha} \\ \quad + (1-\beta)(\lambda + \mathrm{i}\theta)^{\alpha} - 2\lambda^{\alpha}], & \alpha \neq 1, \\ \frac{1}{\pi}\gamma[(1+\beta)(\lambda - \mathrm{i}\theta)\ln(\lambda - \mathrm{i}\theta) \\ \quad + (1-\beta)(\lambda + \mathrm{i}\theta)\ln(\lambda + \mathrm{i}\theta) - 2\lambda\ln\,\lambda], & \alpha = 1, \end{cases} \tag{5.46}$$

with parameter ranges $\alpha \in (0,2)$, $\beta \in [-1,1]$, $\gamma > 0$, $\lambda > 0$ *and* $\delta \in \mathbb{R}$.

The following examples in Barndorff-Nielsen & Shephard (2001) provide interesting special cases of sub-families of the TS distributions presented in Example 5.1 which provide closed-form density expressions.

EXAMPLE 5.1 Closed-Form Density of TS Severity Models

If one considers the loss random variable $X \in \mathbb{R}$ with a TS distribution $X \sim TS(\alpha, \beta = 1, \gamma, \delta, \lambda)$, then its density function is obtained from the tilting of a positive stable distributions density according to

$$\begin{aligned} &f(x; \alpha, \beta = 1, \gamma, \delta = 0, \lambda) \\ &\qquad = f(x; \alpha, \beta = 1, \gamma, \delta = 0; S1)\exp\left(\gamma\lambda - \frac{1}{2}\lambda^{1/\alpha}x\right), \end{aligned} \tag{5.47}$$

for $\alpha \in (0,1)$, $\lambda > 0$, $\gamma \geq 0$ and $\nu \in \mathbb{R}$ with $\lambda \vee (-\nu) > 0$. In the following two special cases, one obtains the closed-form expressions for the densities:

1. If $\alpha = \frac{1}{2}$, then one has a TS density given by an inverses Gaussian severity model (see Chapter 4)

$$\begin{aligned} &f(x; \alpha = 1/2, \beta = 1, \gamma, \delta = 0, \lambda) \\ &\qquad = \frac{\gamma}{\sqrt{2\pi}}\exp(\gamma\lambda)x^{-3/2}\exp\left(-\frac{1}{2}\left(\frac{\gamma^2}{x} + \lambda^2 x\right)\right). \end{aligned}$$

2. If $\alpha = \frac{1}{3}$, then one has a TS density given by

$$\begin{aligned} &f(x; \alpha = 1/3, \beta = 1, \gamma, \delta = 0, \lambda) \\ &\qquad = \frac{\sqrt{2}}{\pi}\gamma^{3/2}\exp\left(\gamma^{3/2}\lambda\right)x^{-3/2}K_{1/3}\left(\left(\frac{2}{3}\gamma\right)^{3/2}x^{-\frac{1}{2}}\right)\exp\left(-\frac{1}{2}\lambda^3 x\right), \end{aligned}$$

with Bessel function $K_{1/3}(x) = \frac{\pi}{\sqrt{3}}\exp(-3x^{1/3})$ and this result is derived in Barndorff-Nielsen & Shephard (2001, example 4.2) using a result from Roberts & Kaufman (1966).

In the case of $\alpha = \frac{1}{2}$, one obtains severity distributions with the following shapes in Figure 5.1, clearly suitable for OpRisk modeling purposes.

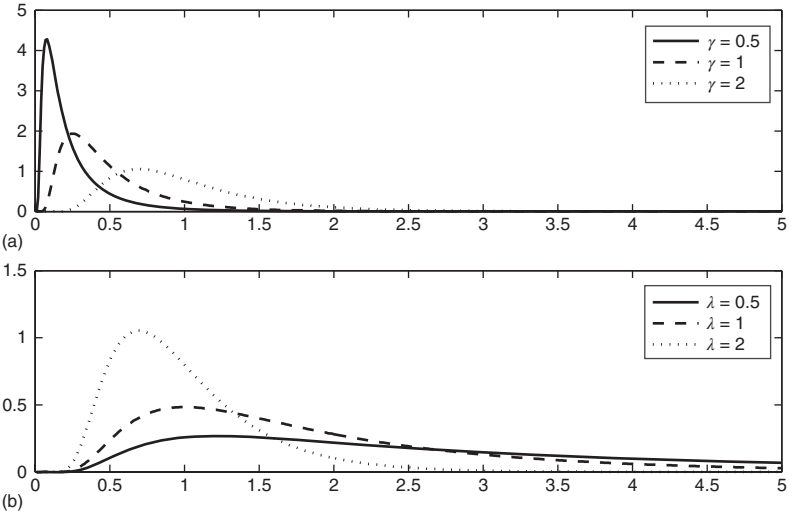

FIGURE 5.1 **Example of tempered stable severity model with $\alpha = 0.5$ for a range of tempering and scale parameters. (a) Tempering parameter $\lambda = 2$ and (b) scale parameter $\gamma = 2$.**

In the case of $\alpha = \frac{1}{3}$, one obtains severity distributions with the following shapes in Figure 5.2, clearly suitable for OpRisk modeling purposes.

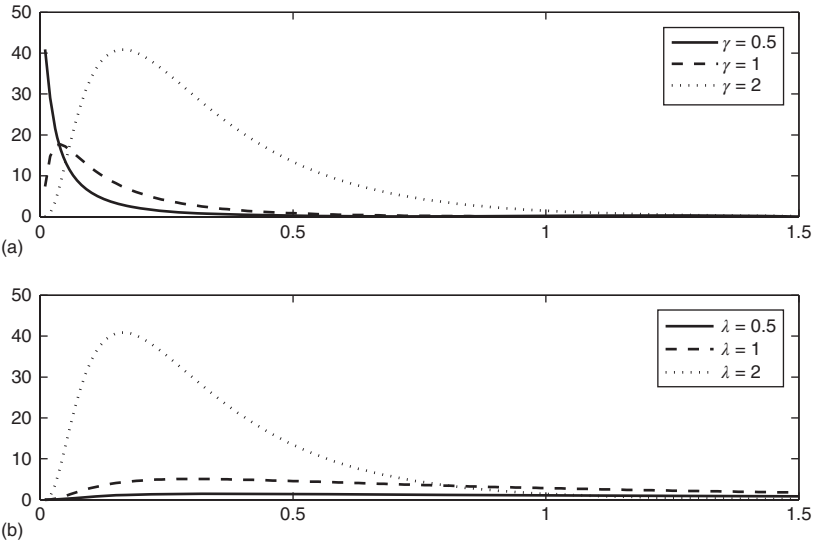

FIGURE 5.2 **Example of tempered stable severity model with $\alpha = 1/3$ for a range of tempering and scale parameters. (a) Tempering parameter $\lambda = 2$ and (b) scale parameter $\gamma = 2$.**

As per the growing popularity of TS processes in financial and actuarial mathematics, there is also a sixth and seventh family of TS models corresponding to the modified stable (MS) and normal modified stable (NMS) processes developed by Barndorff-Nielsen & Shephard (2001). These families of TS distributions are related to other severity model families discussed in Cruz *et al.* (2015). In particular, the MS family are tilted (exponential power tempered) versions of positive stable distributions with $\alpha \in (0,1)$. Then the NMS family is a mixture representation comprised of mixing distributions built from the MS family, analogous to the normal variance–mean mixtures discussed in Chapter 4 when introducing the Normal Inverse Gamma (NIG) severity models.

To construct the MS laws, one considers a positive stable law under say Nolans S1-Type parameterizations with density given by $f(x;\alpha,\beta=1,\gamma,\delta=0;S1)$ for $\alpha \in (0,1)$. Then define the tilted distribution with tilting (tempering) parameter $\lambda > 0$ according to Representation 6 given in Definition 5.9. Note by construction, these MTS severity models have the desirable property that they have strictly positive support.

Definition 5.9 (Representation 6 of Tempered Stable Severity Model (MTS)) *A loss random variable $X \in \mathbb{R}^+$ has an MTS (or MS) distribution $X \sim MS(\alpha,\beta=1,\gamma,\delta,\lambda,\nu;R6)$ if its density function is obtained from the tilting of a positive stable distributions density according to*

$$f(x;\alpha,\beta=1,\gamma,\delta=0,\lambda,\nu;R6) = \frac{x^{\nu+\alpha}}{Z(\alpha,\gamma,\lambda,\nu)} \underbrace{f(x;\alpha,\beta=1,\gamma,\delta=0;S1)\exp\left(\gamma\lambda - \frac{1}{2}\lambda^{1/\alpha}x\right)}_{\text{Tempered stable distribution}}, \quad (5.48)$$

for $\alpha \in (0,1)$, $\lambda > 0$, $\gamma \geq 0$ and $\nu \in \mathbb{R}$ with $\lambda \vee (-\nu) > 0$. The function of parameters $Z(\alpha,\gamma,\lambda,\nu)$ is the normalization constant. The density for the positive perfectly skewed stable distribution is given generically by the series representation for the appropriate parameter settings, given in Lemma 4.7 in the Stable severity model section, see Zolotarev (1986, equation 2.4.6, p. 89) and Feller (2008).

In Bianchi *et al.* (2011, example 3), they state the following results regarding the MTS family of severity models which were parameterized instead via the typical TS form with parameters for the tempering of the Lévy measure given by $c_\pm > 0$ and the tempering factors $\lambda_\pm$. We note that the values $c_\pm$ are the analogues of the C_1 and C_2 that were used by Zolatarev to define the Lévy measure for the α-stable family of distributions. In addition, we note that one can consider that the MTS family is a sub-family of the class of TID severity models.

Definition 5.10 (Lévy Measure of Modified Tempered Stable Severity Models) *A loss random variable $X \in \mathbb{R}^+$ has an MTS (or MS) distribution $X \sim MS(\alpha,\beta=1,\gamma,\delta,\lambda,\nu;R6)$ if its Lévy spectral measure takes the form*

$$W(dx) = \left[c_+ \frac{\lambda_+^{\frac{1}{2}(1+\alpha)} K_{\frac{1}{2}(1+\alpha)}\left(\lambda_+ x\right)}{x^{\frac{1}{2}(1+\alpha)}} \mathbb{I}_{x>0} + c_- \frac{\lambda_-^{\frac{1}{2}(1+\alpha)} K_{\frac{1}{2}(1+\alpha)}(\lambda_- |x|)}{|x|^{\frac{1}{2}(1+\alpha)}} \mathbb{I}_{x<0} \right] dx, \quad (5.49)$$

and the resulting Rosiński measure is given by

$$R(dx) = \left[c_+ \lambda_+^{\alpha+1} \exp\left(-\frac{\lambda_+^2 x^2}{2}\right) \mathbb{I}_{x>0} + c_- \lambda_-^{\alpha+1} \exp\left(-\frac{\lambda_-^2 x^2}{2}\right) \mathbb{I}_{x<0} \right] dx. \quad (5.50)$$

In addition, Bianchi *et al.* (2011) utilize the MTS Lévy measure representation to derive the following result for the characteristic function for the MTS family (as a sub-family of the TID distributions).

Proposition 5.3 (Characteristic Function for the MTS Severity Models) *A loss random variable $X \in \mathbb{R}^+$ has an MTS (or MS) distribution $X \sim MS(\alpha, \delta, c_+, c_-, \lambda_+, \lambda_-; R6)$ if its characteristic function is given by*

$$\Phi_X(\theta) = \exp(\psi(\theta)), \tag{5.51}$$

with

$$\psi(\theta) = \mathrm{i}\theta\delta + c_+ H(\mathrm{i}\theta; \alpha, \lambda_+) + c_- H(-\mathrm{i}\theta; \alpha, \lambda_-), \tag{5.52}$$

and the function

$$\begin{aligned} H(x; \alpha, \lambda) = &\left[\frac{\lambda^\alpha \sqrt{\pi}}{2^{\frac{1}{2}(\alpha+3)}} \Gamma\left(-\frac{\alpha}{2}\right) \left(\left(1 + \frac{x^2}{\lambda^2}\right)^{1/\alpha} - 1 \right) \right] \\ &+ \frac{\mathrm{i}\lambda^{\alpha-1} x}{2^{\frac{1}{2}(\alpha+1)}} \Gamma\left(\frac{1}{2} - \frac{\alpha}{2}\right) \left[2F_1\left(1, \frac{1}{2} - \frac{\alpha}{2}; \frac{3}{2}; -\frac{x^2}{\lambda^2}\right) - 1 \right]. \end{aligned} \tag{5.53}$$

Remark 5.4 (Relationship Between the MTS Family, GIG Family and Lévy Families) *It is interesting to note that the MTS family of severity models, developed by Barndorff-Nielsen & Shephard (2001), have the property that when $\alpha = 1/2$, instead of recovering the Lévy distributional family (as would occur when the tempering was removed with $\lambda \downarrow 0$ and $\nu \to 0$), instead with the tempering one recovers the GIG distributional family studied in Chapter 4.*

To complete the different representations and characterizations of the TS families, we also present the idea of Barndorff-Nielsen & Shephard (2001) which was to create an analogous family of TS distributions using the same mixing ideas as those used to construct the Generalized Hypergeometric families, considered in Cruz *et al.* (2015). The result they considered is to develop the NMTS family as given in Definition 5.11.

Definition 5.11 (Representation 7 of Tempered Stable Severity Model (NMTS)) *A loss random variable $X \in \mathbb{R}^+$ has an NMTS (or NMS) distribution $X \sim NMS(\alpha, \beta = 1, \gamma, \delta, \lambda, \nu, a, b; R7)$ if its density function is obtained from the variance–mean mixture representation given by*

$$X \overset{d}{=} a + bY + \sqrt{Y} Z \tag{5.54}$$

with $Y \sim MS(\alpha, \beta = 1, \gamma, \delta, \lambda, \nu; R6)$, $Z \sim Normal(0, 1)$ and $Y \perp Z$. The parameters satisfy $a \in \mathbb{R}$, $b > 0$, $\alpha \in (0, 1)$, $\lambda > 0$, $\gamma \geq 0$ and $\nu \in \mathbb{R}$ with $\lambda \vee (-\nu) > 0$.

Remark 5.5 *Given these characteristic function representations, the density, distribution functions and quantile functions of each of the TS families can be analogously found using the numerical procedures presented in the α-stable Chapter 4.*

5.1.3 DENSITY OF THE TEMPERED STABLE SEVERITY MODEL

In terms of working with the density of a TS severity model, the approximation considered in Palmer *et al.* (2008) was reported to work well in practical settings and so will be considered later as a suitable approximation. It proceeds from the representation of the TS family of densities given by the tilting of the perfectly skewed α-stable model with $\alpha \in (0, 1)$, which can then be rewritten under the parameterization of Tweedie (1984). This involves the series expansion which was obtained from the Zolatarev series expansions for the perfectly skewed positive stable distributions discussed previously, which is modified for the TS family. In effect, Tweedie considered a class of models with a variance function which was of the form $\mathbb{V}\text{ar}(\mu) \propto \mu^p$ where for some $p > 2$ one obtains the TS family, see related discussions on this family of models in Peters *et al.* (2009c). The resulting TS family density (in the parameterization of Tweedie and Hougard) is then given by

$$\begin{aligned} f(x; \mu, \nu, \alpha) &= -\exp\left(-ax + \frac{a^\alpha b}{\alpha}\right) \frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} \left(-\frac{b}{\alpha x^\alpha}\right)^k \sin(\alpha k \pi) \\ &= \exp\left(-ax + \frac{a^\alpha b}{\alpha}\right) f_X(x; \alpha, \beta = 1, 1, 0; B), \end{aligned} \tag{5.55}$$

with $f_X(x; \alpha, \beta, 1, 0; B)$ the standardized density of the Zolotarev Type B univariate α-Stable model (see Chapter 4), with $\alpha = (p-2)/(p-1)$, $p > 2$ and the parameter settings

$$a = \frac{1-\alpha}{\mu \nu^2}, \quad b = \mu a^{1-\alpha}, \quad a, b > 0. \tag{5.56}$$

It is trivial to see that this corresponds to simply using the series expansion for the α-stable density in Lemma 4.7 with $\alpha < 1$ and $\beta = 1$. In addition, one can see that this corresponds to the parameterization presented in Representation 6, that is, $f(x; \alpha, \beta = 1, \gamma, \delta = 0, \lambda, \nu; R6)$ when one considers the parameters $\mu = \gamma$ and $\nu = \lambda$. There is also a detailed discussion on the Poisson–Tweedie family of closed-form LDA models in Cruz *et al.* (2015, chapter 15) and see Dunn & Smyth (2005).

Proceeding in the representation using parameters (μ, ν, α), one may obtain the Laplace transform of the $X \sim TS(\mu, \nu, \alpha)$ loss random variable written in terms of the the parameter functions in Equation (5.56) and given by

$$f^*(s) = \mathbb{E}[\exp(-sX)] = \exp\left[-\frac{b}{\alpha}\left((a+s)^\alpha - a^\alpha\right)\right], \tag{5.57}$$

which was approximated via a saddle point series expansion in Hougaard (1986) and Palmer *et al.* (2008); they inverted the Laplace transform numerically via the Bromwich inversion integral given in Definition 5.12, see Abate *et al.* (2000).

Definition 5.12 (Bromwich Inversion Integral for Laplace Transforms) *Conisder the Laplace transform of a random variable $X \sim F(x)$ given by*

$$\mathscr{L}[F(x)] = \mathbb{E}[\exp(-sX)] = \int \exp(-sx)\, dF(x), \tag{5.58}$$

or if it admits a density function given by

$$f^*(s) = \mathscr{L}[f(x)] = \mathbb{E}[\exp(-sX)] = \int \exp(-sx) f(x)\, dx. \tag{5.59}$$

Then the inverse Laplace transform to get the resulting density can be obtained by solving the following integration on the complex plane

$$f(x) = \mathscr{L}^{-1}[f^*(s)] = \frac{1}{2\pi \mathrm{i}} \int_{a-\mathrm{i}\infty}^{a+\mathrm{i}\infty} \exp(sx) f^*(s)\, ds, \tag{5.60}$$

where a is a real positive number that is larger than the real component of singularities of $\exp(sx)f^*(s)$ and the integration proceeds along a line at separation a form the origin that is infinite and parallel to the imaginary axis. Note that the integral is zero for $x < 0$.

Note, there is flexibility in choosing the appropriate contour to integrate along, as long as it is to the right of all the singularities of $f^*(s)$. As noted in Abate *et al.* (2000), one can, therefore, select a specification and make a change of variable such that $s = a + \mathrm{i}u$ to obtain

$$f(x) = \mathscr{L}^{-1}[f^*(s)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp((a + \mathrm{i}u)x) f^*(a + \mathrm{i}u)\, du, \tag{5.61}$$

and then utilize the identity

$$\exp((a + \mathrm{i}u)x) = \exp(ax)[\cos(ux) + \mathrm{i}\sin(ux)], \tag{5.62}$$

and the identities $\sin(ux) = -\sin(-ux)$, $\cos(ux) = \cos(-ux)$, $\mathscr{I}\mathrm{m}\{f^*(a + \mathrm{i}u)\} = -\mathscr{I}\mathrm{m}\{f^*(a - \mathrm{i}u)\}$ and $\mathscr{R}\mathrm{e}\{f^*(a + \mathrm{i}u)\} = \mathscr{I}\mathrm{m}\{f^*(a - \mathrm{i}u)\}$ in the Bromwich inversion integral to obtain two integrals that can be solved by quadrature given by

$$\begin{aligned} f(x) &= \frac{2\exp(ax)}{\pi} \int_0^{\infty} \mathscr{R}\mathrm{e}\{f^*(a + u\mathrm{i})\} \cos(ux)\, du, \quad x > 0 \\ &= -\frac{2\exp(ax)}{\pi} \int_0^{\infty} \mathscr{I}\mathrm{m}\{f^*(a + u\mathrm{i})\} \sin(ux)\, du, \quad x > 0. \end{aligned} \tag{5.63}$$

Using this integral transform on the TS densities Laplace transform under a trapezoidal weighting scheme, Palmer *et al.* (2008) obtain the alternating series representation given by

$$f(x; \mu, \nu, \alpha) = \frac{\exp\left(A/\left(2lx\right)\right)}{2} \left[\frac{1}{2} \mathscr{R}\mathrm{e}\left\{f^*(b)\right\} + \sum_{k=1}^{\infty} (-1)^k \mathscr{R}\mathrm{e}\left\{ f^* \left(b + \frac{k\pi}{x} \mathrm{i} \right) \right\} \right] \tag{5.64}$$

with parameter settings for the numerical procedure used given by A, l, m and n, where recommendations are made to use $A = 18.4$, $l = 1$, $m = 12$ and $n = 38$.

5.1.4 PROPERTIES OF TEMPERED STABLE SEVERITY MODELS

In this section, we discuss some useful properties of several of the representations of the TS families of severity models. We begin with general results derived in Rosiński (2007) for TS distributions characterized by Representation 1 of the TS distribution. The first set of results in Proposition 5.4 with respect to fractional powers, the TS analogue of the fractional lower order moment (FLOM) results presented previously for stable severity models.

Proposition 5.4 (Fractional Lower Order Moments of TS Severity Models (FLOMS-TS)) *If X is an infinitely divisible loss random variable with a TS distribution $X \sim TS(\alpha, R(dx), b; R1)$, then the following moment conditions apply, where the distribution is denoted generically here by $F(dx)$, giving:*

$$\begin{aligned}
&\int |x|^p F(dx) < \infty, && p \in (0, \alpha), \\
&\int |x|^\alpha F(dx) < \infty, && \text{if} \quad \int_{|x|>1} |x|^\alpha \ln |x| R(dx) < \infty, \\
&\int |x|^p F(dx) < \infty, && \text{if} \quad \int_{|x|>1} |x|^p R(dx) < \infty, \quad \text{with} \quad p > \alpha, \\
&\int \exp(p|x|) F(dx) < \infty, && \text{if} \quad \int_{|x|>1/p} R(dx) = 0, \quad p > 0.
\end{aligned}$$

As with the class of α-stable severity models, the class of TS severity models are infinitely divisible and self-decomposable; in other words, they satisfy the condition of closure under convolution presented in Proposition 5.5, see Rosiński (2007, corollary 2.12).

Proposition 5.5 (Closure Under Convolution of Tempered Stable Severity Models (R1)) *If $\{X_i\}_{i=1}^n$ are independent infinitely divisible loss random variables each with a TS distribution $X_i \sim TS(\alpha, R_i(dx), b_i; R1)$, then if $\alpha \in (0,1)$ and $\int |x| R(dx) < \infty$, one has*

$$S_n = \sum_{i=1}^n X_i \sim TS\left(\alpha, \sum_{i=1}^n R_i(dx), \sum_{i=1}^n b_i\right). \tag{5.65}$$

These closure properties for TS models used to derive the MTS and NMTS models in Representations 6 and 7 were also discussed in Barndorff-Nielsen & Shephard (2001) to have the property of closure under convolution (infinite divisibility and self-decomposability). If one considers the loss random variable $X \sim TS(\alpha, \beta, \gamma, \delta, \lambda; R6)$ which is built from a strictly positive perfectly skewed stable distribution, then the infinitely divisible loss random variable X is self-decomposable for the following parameter settings:

1. If $\alpha = \frac{1}{2}$, one obtains the GIG laws which are self-decomposable;
2. If $\nu = -\alpha$, one obtains a class of self-decomposable infinitely divisible distributions (closed under convolution);
3. If $-(\alpha + \nu) \in \mathbb{J}^+$, one obtains a class of self-decomposable infinitely divisible distributions (closed under convolution);

4. If $\alpha = 1/m, \quad \forall m = 2, 3, \ldots$ and $\lambda > 0$ with $\alpha + \nu \in \mathbb{J}^+$, then one obtains a class of self-decomposable infinitely divisible distributions (closed under convolution).

In addition, provided that the MS family is infinitely divisible then the same properties discussed earlier will also apply to the NMS family of severity models, see discussion in Barndorff-Nielsen & Shephard (2001). The properties of closure under convolution then apply to the following models as given in Definition 5.13.

Definition 5.13 (Closure Under Convolution Properties of TS Severity Models) *If* $\{X_i\}_{i=1}^n$ *are independent infinitely divisible loss random variables each with a TS distribution* $X_i \sim TS(\alpha, R_i(dx), b_i; R1)$*, then if* $\alpha \in (0, 1)$ *and* $\int |x| R(dx) < \infty$*, one has*

$$S_n = \sum_{i=1}^{n} X_i \sim TS\left(\alpha, \sum_{i=1}^{n} R_i(dx), \sum_{i=1}^{n} b_i\right). \tag{5.66}$$

The following results apply

- $X_i \sim CTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim CTS(\alpha, nc, \lambda_+, \lambda_-, nm)$.
- $X_i \sim RDTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim RDTS(\alpha, nc, \lambda_+, \lambda_-, nm)$.
- $X_i \sim GTS(\alpha_+, \alpha_-, c_+, c_-, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim CTS(\alpha_+, \alpha_-, nc_+, nc_-, \lambda_+, \lambda_-, nm)$.
- $X_i \sim MTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim MTS(\alpha, nc, \lambda_+, \lambda_-, nm)$.
- $X_i \sim NMTS(\alpha, c, \lambda, b, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim NMTS(\alpha, nc, \lambda, b, nm)$.
- $X_i \sim KRTS(\alpha, k_+, k_-, r_+, r_-, p_+, p_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim KRTS(\alpha, nk_+, nk_-, r_+, r_-, p_+, p_-, nm)$.

In addition, one can state the cumulants for this characteristic function of the CTS model which are given in Proposition 5.6.

Proposition 5.6 (Cumulants of CTS Severity Models) *A loss random variable* X *which is distributed according to a CTS model* $X \sim CTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R2)$ *has cumulants given by the following expression*

$$\begin{aligned}\kappa_n[X] &= \frac{1}{\mathrm{i}^n} \frac{d^n}{d\theta^n} \ln \mathbb{E}\left[\exp(\mathrm{i}\theta X)\right]\Big|_{\theta=0} \\ &= \begin{cases} m, & n = 1, \\ \Gamma(n-\alpha)\left[c_+ \lambda_+^{\alpha-n} + (-1)^n c_- \lambda_-^{\alpha-n}\right], & n = 2, 3, \ldots \end{cases}\end{aligned} \tag{5.67}$$

One can then obtain a standardized loss random variable by recognizing the result in Proposition 5.7.

Proposition 5.7 (Standardized CTS Severity Models) *A loss random variable X which is distributed according to a CTS model $X \sim CTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R2)$ is standardized in the sense that it has zero mean and unit variance if one selects*

$$c = c_+ = c_- = \frac{1}{\Gamma(2-\alpha)[\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}]}. \tag{5.68}$$

The cumulants of the RDTS sub-family of TS severity models are also available in closed form according to the expression given in Proposition 5.14, see Shin Kim *et al.* (2010).

Definition 5.14 (Cumulants of the RDTS Severity Model) *An infinitely divisible loss random variable X with an RDTS distribution $X \sim RDTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R3)$ has cumulants given by*

$$\begin{aligned}\kappa_n[X] &= \frac{1}{\mathrm{i}^n}\frac{d^n}{d\theta^n}\ln \mathbb{E}\left[\exp(\mathrm{i}\theta X)\right]\Big|_{\theta=0} \\ &= \begin{cases} m, & n=1 \\ 2^{\frac{1}{2}(n-\alpha-2)}\Gamma((n-\alpha)/2)\left[c_+\lambda_+^{\alpha-n} + (-1)^n c_-\lambda_-^{\alpha-n}\right], & n=2,3,\ldots \end{cases}\end{aligned} \tag{5.69}$$

As with the CTS family of severity distributions, one may also define the standardized severity distribution by the result in Proposition 5.8.

Proposition 5.8 (Standardized RDTS Severity Models) *A loss random variable X which is distributed according to an RDTS model $X \sim RDTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R3)$ is standardized in the sense that it has zero mean and unit variance if one selects*

$$c = c_+ = c_- = \frac{2^{\alpha/2}}{\Gamma\left(1-\alpha/2\right)\left[\lambda_+^{\alpha-2} + \lambda_-^{\alpha-2}\right]}. \tag{5.70}$$

One can also obtain cumulant-generating functions for the GTS, MTS and Kim-Rachev tempered stable (KRTS) families of severity models as presented in Definitions 5.15–5.17.

Definition 5.15 (Cumulants of the GTS Severity Model) *An infinitely divisible loss random variable X with a GTS distribution $X \sim GTS(\alpha_+, \alpha_-, c_+, c_-, \lambda_+, \lambda_-, m; R4)$ has cumulants given by*

$$\begin{aligned}\kappa_n[X] &= \frac{1}{\mathrm{i}^n}\frac{d^n}{d\theta^n}\ln \mathbb{E}\left[\exp(\mathrm{i}\theta X)\right]\Big|_{\theta=0} \\ &= \begin{cases} m, & n=1, \\ c_+\Gamma\left(n-\alpha_+\right)\lambda_+^{\alpha_+-n} + (-1)^n c_-\Gamma(n-\alpha_-)\lambda_-^{\alpha_--n}, & n=2,3,\ldots \end{cases}\end{aligned} \tag{5.71}$$

Definition 5.16 (Cumulants of the MTS Severity Model) *An infinitely divisible loss random variable X with an MTS distribution $X \sim MTS(\alpha, c, \lambda_+, \lambda_-, m; R6)$ has cumulants given by*

$$\kappa_n[X] = \frac{1}{\mathrm{i}^n}\frac{d^n}{d\theta^n}\ln \mathbb{E}\left[\exp(\mathrm{i}\theta X)\right]\Big|_{\theta=0}$$
$$= \begin{cases} m, & n=1 \\ 2^{n-\frac{1}{2}(\alpha+3)}\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{n-\alpha}{2}\right)[\lambda_+^{\alpha-n}+(-1)^n\lambda_-^{\alpha-n}], & n=2,3,\ldots \end{cases}$$

Definition 5.17 (Cumulants of the KRTS Severity Model) *An infinitely divisible loss random variable X with a KRTS distribution $X \sim MRTS(\alpha, k_+, k_-, r_+, r_-, p_+, p_-, \delta; R6)$ has cumulants given by*

$$\kappa_n[X] = \frac{1}{\mathrm{i}^n}\frac{d^n}{d\theta^n}\ln \mathbb{E}\left[\exp(\mathrm{i}\theta X)\right]\Big|_{\theta=0}$$
$$= \begin{cases} \delta, & n=1, \\ \Gamma(n-\alpha)\left[\frac{k_+ r_+^n}{p_+ + n} + (-1)^n \frac{k_- r_-^n}{p_- + n}\right], & n=2,3,\ldots \end{cases}$$

In representing such TS severity models, one can again utilize ideas developed previously in the α-stable families properties. This was the approach adopted to develop the main result in Poirot & Tankov (2006).

Remark 5.6 *The work of Poirot & Tankov (2006) utilizes an analogous idea for TS processes as presented in the stable severity model setting in Proposition 4.3 where any stable severity model was shown to be decomposed according to a linear combination of two suitable scaled (one reflected) perfectly skewed stable distributions. It is clearly, therefore, the case that one may obtain such a representation also for the TS severity models.*

Remark 5.7 (Estimation of Tempered Stable Severity Model Parameters) *Regarding parameter estimation for any of these models, because the cumulant-generating functions are available, one may perform parameter estimation through moment-matching-based methods or alternative via numerical MLE using the density representations discussed previously.*

5.1.5 PARAMETER ESTIMATION OF LOSS RANDOM VARIABLES FROM A TEMPERED STABLE SEVERITY MODEL

There has been little work discussing the parameter estimation for TS models. The only paper we are aware of that discusses this issue is the work of Palmer *et al.* (2008) where a numerical maximum-likelihood estimation (MLE) procedure is utilized via the density approximation presented previously. In this section, we discuss parameter estimation via moment matching as detailed in Proposition 5.9.

Proposition 5.9 (Method of Moments for Perfectly Skewed RDTS Severity Models) *Consider a loss random variable X with RDTS distribution*

$X \sim RDTS(\alpha, c_+, c_-, \lambda_+, \lambda_-, m; R3)$ such that one has n i.i.d. loss random variables that are observed in a given year $\{X_i\}_{i=1}^n$. Then the sample mean, variance, skewness and excess kurtosis are denoted by

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2,$$

$$\widehat{S} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2\right)^{\frac{3}{2}}},$$

$$\widehat{K}_e = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^4}{\frac{1}{n}\left(\sum_{i=1}^{n}(X_i - \widehat{\mu})^2\right)^2} - 3.$$

Then given these sample estimates, it would be typical in OpRisk modeling to expect $\lambda_+ > \lambda_-$, and in the perfectly skewed case, one sets $\lambda_- = 0$ and estimation proceeds for α, c, λ_+ via the population moments matched to the sample moments. First, one observes that the estimator of m is given by the sample mean $\widehat{m} = \widehat{\mu}$. Then the following systems of equations are solved:

$$\widehat{\sigma}^2 = 2^{-\alpha/2}\Gamma\left(1 - \frac{\alpha}{2}\right) c\lambda_+^{\alpha-2},$$

$$\widehat{S} = 2^{\frac{1}{4}(\alpha+2)} \frac{\Gamma\left(\frac{3-\alpha}{2}\right) c\lambda_+^{\alpha-3}}{\left(\Gamma\left(1 - \frac{\alpha}{2}\right) c\lambda_+^{\alpha-2}\right)^{\frac{3}{2}}},$$

$$\widehat{K}_e = 2^{1+\alpha/2} \frac{\Gamma\left(\frac{4-\alpha}{2}\right) c\lambda_+^{\alpha-4}}{\left(\Gamma\left(1 - \frac{\alpha}{2}\right) c\lambda_+^{\alpha-2}\right)^2}.$$

This system of equations is then solved numerically via a Newton method where if λ_- had been included one would use all three equations, however to obtain the estimators for the model parameters λ_+, and α under the assumption of perfect skew, i.e. that $\lambda_- = 0$, then one can select the first two equations to solve.

One may also employ a more generalized approach to the parameter estimation based on the known forms for the population cumulants and the central moments sample estimators as detailed in Proposition 5.10. This procedure is a form of generalized method of Moments.

Proposition 5.10 (Parameter Estimation of TS Models via Generalized Cumulant Matching) *The population cumulants of any of the TS models (denoted by $\kappa_n[X]$) are related to the population moments via the following recursion*

$$\kappa_n[X] = \mathbb{E}[X^n] - \sum_{m=1}^{n-1} \frac{(n-1)!}{(n-m-2)!(m-1)!}\kappa_m[X]\mathbb{E}[X^{n-m}]. \tag{5.72}$$

In the case of the TS models, the largest number of parameters is in the GTS or KRTS models which have eight parameters meaning they will require at least this many cumulant equations for a system

of equations to solve the parameters, the first few cumulants are given by

$$\kappa_1[X] = \mathbb{E}[X]$$
$$\kappa_2[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
$$\kappa_3[X] = \mathbb{E}[X^3] - 3\mathbb{E}[X^2]\mathbb{E}[X] + 2\mathbb{E}[X]^2$$
$$\kappa_4[X] = \mathbb{E}[X^4] - 4\mathbb{E}[X^3]\mathbb{E}[X] - 3\mathbb{E}[X^2]^2 + 12\mathbb{E}[X^2]\mathbb{E}[X]^2 - 6\mathbb{E}[X]^4.$$

Note that the coefficients of the cumulants follow those of the Bell polynomial sequence. To perform the estimation, one then estimates the sample non-central moments and obtains a system of equations using the known population cumulants, which are solved numerically via Newton methods.

Remark 5.8 *As the sample moments are asymptotically consistent estimators of the population moments as the number of loss samples $n \to \infty$, then one would expect that under the cumulant-matching procedure above, the estimated parameters for the TS model should also be consistent estimators.*

5.1.6 SIMULATION OF LOSS RANDOM VARIABLES FROM A TEMPERED STABLE SEVERITY MODEL

There have been numerous discussions in the literature about sampling of TS loss random variables. For instance, there are the approaches of Brix (1999), Rosiński (2001), the Laplace transform approach of Ridout (2009) and the characteristic function rejection approaches of Devroye (2009). There has also been a discussion and summary of these approaches in Hofert (2011). The basic approaches include:

1. The series representation approach with deterministic or stochastic truncations, see discussions in Rosiński (2001) and Imai & Kawai (2011). The finite truncation effects of these infinite shot noise series representations is studied with regard to the introduced bias;
2. The inverse Laplace transform methods of Ridout (2009) for sampling exponential tilted stable distributions;
3. The generalized rejection envelope method developed for characteristic functions adapted to the case of non-negative random variables in the work of Devroye (1981) and the recent updated ideas in Devroye (2009).

Note, the exact sampling approach using numerical inversion and rejection proposed in Ridout Ridout (2009) and later also discussed by Hofert (2011) was itself independently it seems discovered earlier by Kawai & Masuda (2009) and later in Kawai & Masuda (2011).

We start this discussion by presenting the shot noise series representations of the TS families of loss random variables that will lead to some simulation strategies. If one considers Representations 1 and 2 of the TS severity model given by the CTS and RDTS families, then the following convergent series representations in Theorem 5.2 are available which may be truncated, either stochastically or deterministically to provide a means of simulation from such severity models, see discussion in Rosiński (2007), Shin Kim *et al.* (2010) and Bianchi *et al.* (2011). These series expansions are the direct analogues of those obtained by Zolatarev and presented in the previous section on representation of Stable severity models. There are also series representations available for the CGMY severity model, the KR severity models discussed in Bianchi *et al.* (2011).

Theorem 5.2 (Convergent Series Shot Noise Representations of TS Severity Models) *Consider the i.i.d. sequence of random variables* $\{U_j\}$ *with* $U_j \sim Uniform(0,1)$, $\{E_j\}$, $\{E'_j\}$ *with* $E_j \sim Exp(1)$ *and* $E'_j \sim Exp(1)$ *and* $\{\nu_j\}$ *given by probability mass function*

$$\mathbb{P}\mathrm{r}[\nu_j = x] = \begin{cases} \frac{1}{2}, & \text{if } \nu_j = \lambda_+, \\ \\ \frac{1}{2}, & \text{if } \nu_j = \lambda_-. \end{cases}$$

Using these independent random sequences, one can develop series representations of the different TS loss random variables as follows:

1. *A loss random variable* X *with CTS severity model* $X \sim CTS(\alpha, c, c, \lambda_+, \lambda_-, 0)$ *can be represented by the almost surely (a.s.) convergent series given by*

$$X \overset{d}{=} \begin{cases} \sum_{j=1}^{\infty} \left[\left(\frac{\alpha}{c} \sum_{k=1}^{j} E'_k \right)^{-1/\alpha} \wedge E_j U_j^{1/\alpha} |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - \Gamma(1-\alpha) c (\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}), \\ \qquad\qquad \alpha \in (0,2) \notin \{0\}, \\ \sum_{j=1}^{\infty} \left[\left(\frac{\alpha}{c} \sum_{k=1}^{j} E'_k \right)^{-1} \wedge E_j U_j |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - c \ln\left(\lambda_+ / \lambda_-\right), \\ \qquad\qquad \alpha = 1. \end{cases}$$

2. *A loss random variable* X *with RDTS severity model* $X \sim RDTS(\alpha, c, c, \lambda_+, \lambda_-, 0)$ *can be represented by the a.s. convergent series given by*

$$X \overset{d}{=} \sum_{j=1}^{\infty} \left[\left(\frac{\alpha}{c} \sum_{k=1}^{j} E'_k \right)^{-1/\alpha} \wedge \sqrt{2} E_j^{1/2} U_j^{1/\alpha} |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|}$$
$$-2^{-\frac{\alpha+1}{2}} c \Gamma\left(\frac{1-\alpha}{2} \right) (\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}), \quad \alpha \in (0,2) \notin \{0\}.$$

3. *A loss random variable* X *with CGMY severity model* $X \sim CGMY(C, G, M, Y, 0)$ *can be represented by the a.s. convergent series given by the definition of* $\{\nu_j\}$ *given by probability mass function*

$$\mathbb{P}\mathrm{r}[\nu_j = x] = \begin{cases} \frac{1}{2}, & \text{if } \nu_j = -G, \\ \\ \frac{1}{2}, & \text{if } \nu_j = M, \end{cases}$$

$$X \overset{d}{=} \begin{cases} \sum_{j=1}^{\infty} \left[\left(\frac{Y}{2C} \sum_{k=1}^{j} E'_k \right)^{-\frac{1}{Y}} \wedge E_j U_j^{\frac{1}{Y}} |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - \Gamma(1-Y) C (M^{Y-1} - G^{Y-1}), \\ \qquad\qquad \alpha \in (0,2) \notin \{0\}, \\ \sum_{j=1}^{\infty} \left[\left(\frac{1}{c} \sum_{k=1}^{j} E'_k \right)^{-1} \wedge E_j U_j |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - (2\mathscr{E} - C(\ln G - \ln M), \\ \qquad\qquad Y = 1, \end{cases}$$

with $\mathscr{E}$ *the Euler constant.*

4. *A loss random variable X with KR severity model $X \sim KR(\alpha, k_+, k_-, r_+, r_-, 0)$ can be represented by the a.s. convergent series given for one of the following two cases:*

- $\alpha \in (0,1)$*; or*
- $\alpha \in [1,2)$ *with* $k_+ = k_-$, $r_+ = r_-$ *and* $p_+ = p_-$.

Then with the definition of $\{\nu_j\}$ *being i.i.d. random variables with distribution*

$$f_\nu(x) = \frac{1}{\eta}\left(k_+ r_+^{-p_+} x^{-\alpha-p_+-1}\mathbb{I}_{x>\frac{1}{r_+}} + k_- r_-^{-p_-}|x|^{-\alpha-p_+-1}\mathbb{I}_{x<-\frac{1}{r_-}}\right)$$

with η *defined by*

$$\eta = \frac{k_+ r_+^\alpha}{\alpha+p_+} + \frac{k_- r_-^\alpha}{\alpha+p_-},$$

one gets the shot noise series representations

$$X \overset{d}{=} \sum_{j=1}^{\infty}\left[\left(\frac{\alpha}{\eta}\sum_{k=1}^{j} E_k'\right)^{-1/\alpha} \wedge E_j U_j^{\frac{1}{\alpha}}|\nu_j|^{-1}\right]\frac{\nu_j}{|\nu_j|} - \Gamma(1-\alpha)\left(\frac{k_+ r_+}{1+p_+} - \frac{k_- r_-}{1+p_-}\right).$$

If $\alpha \in [1,2)$*, then the series representation, with the following definitions*

$$x_0 = \frac{1}{\eta}\left(\frac{k_+ r_+^\alpha}{\alpha+p_+} - \frac{k_- r_-^\alpha}{\alpha+p_-}\right),$$

$$x_1 = \frac{k_+ r_+}{1+p_+} - \frac{k_- r_-}{1+p_-},$$

one gets the series representation

$$X \overset{d}{=} \sum_{j=1}^{\infty}\left[\left(\frac{\alpha}{\eta}\sum_{k=1}^{j} E_k'\right)^{-1/\alpha} \wedge E_j U_j^{\frac{1}{\alpha}}|\nu_j|^{-1}\right]\frac{\nu_j}{|\nu_j|} - \left(\frac{\alpha j}{\eta}\right)^{-1/\alpha} x_0 + b,$$

with

$$b = \begin{cases} \alpha^{-1/\alpha}\zeta_{\frac{1}{\alpha}}\eta^{1/\alpha}x_0 - \Gamma(1-\alpha)x_1, & \alpha \in (1,2), \\ \left(2\mathscr{E} + \ln(\eta)\right)x_1 - \left[\frac{k_+ r_+}{p_+ +1}\left(\ln r_+ - \frac{1}{p_+ +1}\right) - \frac{k_- r_-}{p_- +1}\left(\ln r_- - \frac{1}{p_- +1}\right)\right], & \\ & \alpha = 1, \end{cases}$$

with ζ *the Riemann zeta function.*

Hence, simulation of the CTS stable loss random variable would proceed as follows; clearly given the series representations discussed, these can then be generated for any other TS families. The algorithm presented later will have a truncation bias as detailed in Imai & Kawai (2011).

Algorithm 5.1 (Approximate Simulation of Losses from $CTS(\alpha, c, c, \lambda_+, \lambda_-, 0)$ Severity Models)

1. *Fix a truncation order (based on truncation bias details in Imai & Kawai (2011)) to obtain n.*
2. *Draw i.i.d. uniform random variables $\{U_j\}_{j=1}^n$ such that $U_j \sim Uniform(0,1)$.*
3. *Draw i.i.d. exponential random variables $\{E_j\}_{j=1}^n$ such that $E_j \sim Exp(1)$.*
4. *Draw i.i.d. exponential random variables $\{E'_j\}_{j=1}^n$ such that $E'_j \sim Exp(1)$.*
5. *Draw i.i.d. random variables $\{\nu_j\}_{j=1}^n$ such that*

$$\mathbb{P}\mathrm{r}[\nu_j = x] = \begin{cases} \frac{1}{2}, & \text{if} \quad \nu_j = \lambda_+ \\ \frac{1}{2}, & \text{if} \quad \nu_j = \lambda_-. \end{cases}$$

6. *Construct summation of terms according to the truncated series representation*

$$X \overset{d}{=} \begin{cases} \sum_{j=1}^{n} \left[\left(\frac{\alpha}{c} \sum_{k=1}^{j} E'_k \right)^{-1/\alpha} \wedge E_j U_j^{1/\alpha} |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - \Gamma(1-\alpha) c (\lambda_+^{\alpha-1} - \lambda_-^{\alpha-1}), \\ \qquad\qquad \alpha \in (0,2) \notin \{0\}, \\ \sum_{j=1}^{n} \left[\left(\frac{\alpha}{c} \sum_{k=1}^{j} E'_k \right)^{-1} \wedge E_j U_j |\nu_j|^{-1} \right] \frac{\nu_j}{|\nu_j|} - c \ln\left(\frac{\lambda_+}{\lambda_-} \right), \\ \qquad\qquad \alpha = 1. \end{cases}$$

There are also a number of rejection algorithms and inverse Laplace transform approaches one could consider that are exact methods although they may have a high computational cost because of the numerical inversion and rejection stages. We highlight two basic approaches that are numerically robust and exact to a known order of approximation error because of the numerical inversion of an integral representation. The first approach is the cumulative distribution function (cdf) inversion method.

Algorithm 5.2 (Generic Simulation of Losses from TS Severity Models)

1. *Consider a TS Stable model with generic distribution function $F(x)$, select an interval $[x_{\min}, x_{\max}]$ and a grid spacing given by $x_k = a + hk$ with $h = \frac{b-a}{N}$ for N points and $k \in \{1, 2, \ldots, N\}$.*
2. *Evaluate the distribution function (numerical integral inversion) at points $\{x_k\}_{k=1}^N$ to get $\{F(x_k)\}_{k=1}^N$.*
   (a) Evaluate the trapezoidal density series expansion derived previously at points x_k via

$$f(x_k; \mu, \nu, \alpha) = \frac{\exp\left(\frac{A}{2lx_k}\right)}{2} \left\{ \frac{1}{2} \exp\left[-\frac{b}{\alpha} \left((a+b)^\alpha - a^\alpha \right) \right] + \sum_{k=1}^{\infty} (-1)^k \mathscr{R}\mathrm{e} \left\{ \exp\left(-\frac{b}{\alpha} \left[\left(a + b + \frac{k\pi}{x_k} \mathrm{i} \right)^\alpha - a^\alpha \right] \right) \right\} \right\},$$

where α is the index of stability in any of the specified TS models.

(b) Construct the distribution function using the density via $F(x_{k+1}) = F(x_k) + hf(x_k)$ *with boundary condition* $F(x_0) = F(x_{\min}) = 0$.

3. *Draw a uniform random variable* $U \sim Uniform(0,1)$.
4. *Find index* n *such that* $F(x_n) \leq U < F(x_{n+1})$.
5. *Return draw* $X = x_n + \frac{h[U-F(x_n)]}{F(x_{n+1})-F(x_n)} + m$, *where* m *is the location parameter for the stable model.*

For more computationally efficient and advanced rejection methods, one can consider the rejection squeeze method of Devroye (1981) which uses the characteristic function and its second derivative. Alternatively, there is also the recent work which avoids any numerical integral inversions see Devroye (2009).

5.1.7 TAIL BEHAVIOUR OF THE TEMPERED STABLE SEVERITY MODEL

In Barndorff-Nielsen & Shephard (2001), it is shown that TS severity models admit semi-heavy tails and that in fact the limiting behaviour of the MTS and NMTS models is directly linked to the tail behaviour of the TS severity models. Therefore, we will focus on the TS tail behaviours, which in turn may be derived from the tail behaviour of the tempered series expansions of the perfectly skewed stable models with $\alpha \in (0,1)$. This allows one to obtain the following asymptotic tail behaviour of the TS severity distribution given in Proposition 5.11.

Proposition 5.11 (Tail Behaviour of Tempered Stable Severity Models) *Consider a loss random variable* X *with TS severity model* $X \sim TS(\alpha, \beta = 1, \gamma, \delta = 0, \lambda, \nu; R6)$ *given by the tilting of the positive perfectly skewed stable density with* $\alpha \in (0,1)$ *according to*

$$f(x;\alpha,\beta=1,\gamma,\delta=0,\lambda,\nu;R6) = \underbrace{f(x;\alpha,\beta=1,\gamma,\delta=0;S1)}_{\text{Skewed stable Nolan Type S1}} \exp\left(\gamma\lambda - \frac{1}{2}\lambda^{1/\alpha}x\right).$$

Then using the series expansion for this α*-stable model, one obtains the asymptotic tail behaviours as* $x \to \pm\infty$ *given by*

$$f(x;\alpha,\beta=1,\gamma,\delta=0,\lambda,\nu;R6) \sim \gamma 2^{\alpha}\frac{\Gamma(1+\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}x^{-\alpha-1}.$$

In addition, one can derive the semi-heavy tail behaviour of the NMTS family as shown in Proposition 5.12.

Proposition 5.12 (Tail Behaviour of Normal Modified Tempered Stable Severity Models) *Consider a loss random variable* X *with TS severity model* $X \sim NMS(\alpha, \beta = 1, \gamma, \delta, \lambda, \nu, a, b; R7)$. *Then the asymptotic tail behaviour is given by, as* $x \to \pm\infty$,

$$f(x;\alpha,\beta=1,\gamma,\delta=0,\lambda,\nu;R6)$$
$$\sim 2^{\alpha+1}\gamma e^{\gamma\lambda}\frac{\Gamma(1+\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}\left(a^2+\lambda^{1/\alpha}\right)^{\frac{1}{4}(1+2\alpha)}|x|^{-\alpha-1}\exp\left[-\left(a^2+\lambda^{1/\alpha}\right)^{\frac{1}{2}}|x|+ax\right].$$

5.2 Quantile Function Heavy-Tailed Severity Models

In this section, we discuss a popular distributional family for severity models in OpRisk which can only be specified via the transformation of another standard random variable such as a Gaussian. Examples of OpRisk severity models defined through their quantile functions include Johnson family with base distribution given by Gaussian or logistic and the Tukey family with base distribution Gaussian or logistic. The concept of constructing skew and heavy-tailed distributions through the use of a transformation of a Gaussian random variable was originally proposed in the work of Tukey (1977a) and is, therefore, aptly named the family of Tukey distributions. This family of distributions was then extended by Hoaglin (2006) and Jorge & Boris (1984). Within this family of distributions, two particular sub-families have received the most attention in the literature; these correspond to the g-and-h and the g-and-k distributions. The first of these families have been studied in a few contexts in OpRisk, see Dutta & Perry (2006), Peters & Sisson (2006), Degen *et al.* (2007) and Jiménez & Arunachalam (2011) and the references therein for applications.

Before presenting details of the g-and-h and the g-and-k distributions for an OpRisk audience, we first discuss the general family of Tukey distributions. Basically, Tukey suggested several non-linear transformations of a standard Gaussian random variable, denoted here by $W \sim Normal(0,1)$ so as not to be confused with the annual loss that we denote throughout by Z. The g-and-h transformations involved a skewness transformation g and a kurtosis transformation h. If the kurtosis transformation of the type h is replaced with the type k, one obtains the g-and-k family of distributions discussed in Rayner & MacGillivray (2002). If the h transformation is replaced by the j transformation, one obtains the g-and-j transformations of Fischer & Klein (2004).

We begin with the generic specification of the Tukey transformation given in Definition 5.18. These types of transformations were labelled elongation transformations, where the notion of elongation was noted to be closely related to tail properties such as heavy tailedness, see discussions in Hoaglin (2006). In considering such a class of elongation transformations to obtain a distribution, one is comparing the tail strength of the new distribution with that of the base distribution (such as a Gaussian or logistic). In this regard, one can think of tail strength or heavy tailedness as an absolute concept, where the notion of elongation strength is a relative concept. In the following, we will first consider relative elongation compared to a base distribution for a generic random variable W. It should be clear that such a measure of relative tail behaviour is independent of location and scale. Other properties that such an elongation transform $T(\cdot)$ should satisfy are that it should preserve symmetry $T(w) = T(-w)$; the base distribution should not be significantly transformed in the centre such that $T(w) = w + O(w^2)$ for w around the mode. Then to increase the tails of the resulting distribution, relative to the base, it is important to assume that T is strictly monotonically increasing transform that is convex, that is, for positive $w > 0$, such that $T'(w) > 0$ and $T''(w) > 0$ for $w > 0$. One such transformation family satisfying these properties includes the Tukey Transformations.

Definition 5.18 (Tukey Transformations) *Consider a Gaussian random variable $W \sim Normal(0,1)$ and a suitable transformation $T(w)$ which will produce a new random variable X with Tukey distribution, given by*

$$X = WT(W)^{\theta}, \tag{5.73}$$

for a parameter $\theta \in \mathbb{R}$.

Typically, in the OpRisk setting, it will be desirable when working with such severity models to enforce a constraint that the tails of the resulting distribution after transformation are heavier than the Gaussian distribution. In this case, one should consider a transformation $T(w)$ which is positive, symmetric and strictly monotonically increasing for positive values of $W \geq 0$. In addition, it will be desirable to obtain this property of heavy tails relative to the Gaussian to also consider setting the parameter $\theta \geq 0$. As discussed, a series of kurtosis transformations were proposed in the literature. The Tukey h, k and j transforms are provided in Definition 5.19.

Definition 5.19 (Tukey's Kurtosis Transformations—h, k and j Types) *The h-type of transformation, denoted by $T_h(w)$, is given by*

$$T_h(w) = \exp(w^2). \tag{5.74}$$

The k-type of transformation, denoted by $T_k(w)$, is given by

$$T_k(w) = 1 + w^2. \tag{5.75}$$

The j-type of transformation, denoted by $T_j(w)$, is given by

$$T_j(w) = \frac{1}{2}[\exp(w) + \exp(-w)]. \tag{5.76}$$

EXAMPLE 5.2 Tukey Base Elongation Transforms for Kurtosis: h, j, k

In Figure 5.3, we plot the simple base transforms for the h, j and k Tukey elongation transforms. These plots demonstrate that the base j and k transforms have a similar effect on the tails of the base distribution, which is distinct in the kurtosis introduced by the h transform.

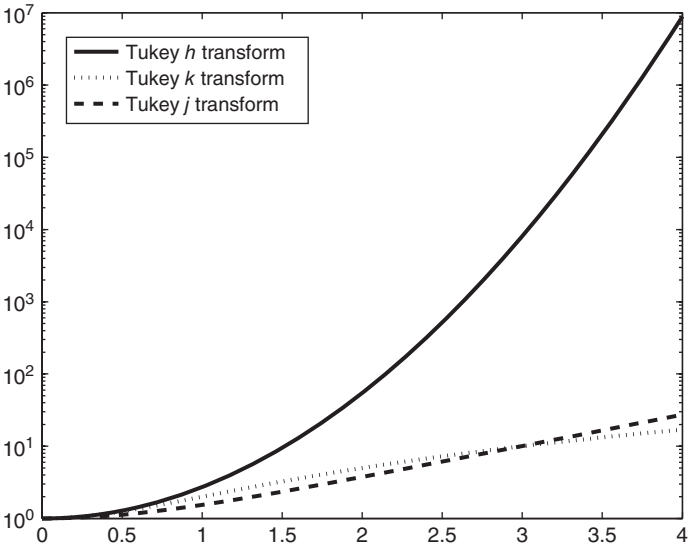

FIGURE 5.3 Base transforms for the Tukey elongation kurtosis transforms.

To nest all these transformations within one class of transformations, the work of Fischer (2010) proposed a power series representation denoted by the subscript 'a' given in Equation 5.77. This suggestion, although it nested the other families of distributions, is not practical for use as it involves the requirement of estimating a very large (infinite) number of parameters a_i to obtain the data-generating mechanism.

$$T_a(w) = \sum_{i=0}^{\infty} a_i w^{2i}. \tag{5.77}$$

As a consequence, this nesting structure was replaced with the general transformation given by Fischer (2010) which took the form given in Equation 5.78

$$T_{hjk}(w;\alpha,\beta,\gamma) = \left(1 + \frac{\left(w^2+\gamma\right)^{\alpha} - \gamma^{\alpha}}{\beta}\right)^{\beta}, \qquad \alpha > 0, \beta \geq 1, \gamma > 0. \tag{5.78}$$

Then it is clear that the original h, k and j transformations are recovered with $T_h(w) = T_{hjk}(w;1,\infty,\gamma)$, $T_k(w) = T_{hjk}(w;1,1,\gamma)$ and $T_j(w) \approx T_{hjk}(w;0.5,\infty,0.5)$.

EXAMPLE 5.3 Tukey Elongation Transform Density Shapes

Consider the Tukey elongation transform of a base reference random variable $W \sim Normal(0,1)$ given by the generic transform superclass of Fischer (2010) according to

$$T_{hjk}(w;\alpha,\beta,\gamma) = \left(1 + \frac{\left(w^2+\gamma\right)^{\alpha} - \gamma^{\alpha}}{\beta}\right)^{\beta}, \qquad \alpha > 0, \beta \geq 1, \gamma > 0.$$

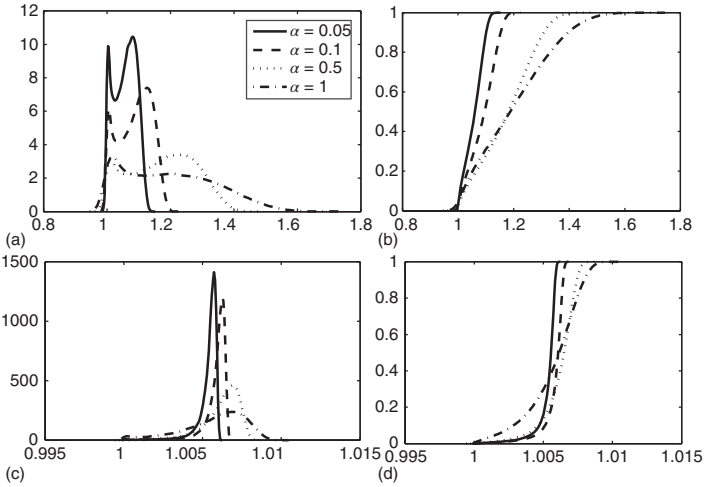

FIGURE 5.4 **(a) The density of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha \in \{0.05, 0.1, 0.5, 1\}$, $\beta = 0.1$ and $\gamma = 0.1$. (b) The distribution function of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha \in \{0.05, 0.1, 0.5, 1\}$, $\beta = 0.1$ and $\gamma = 0.1$. (c) The density of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha \in \{0.05, 0.1, 0.5, 1\}$, $\beta = 0.001$ and $\gamma = 0.001$. (d) The distribution of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha \in \{0.05, 0.1, 0.5, 1\}$, $\beta = 0.001$ and $\gamma = 0.001$.**

Figure 5.4 shows the distributions of this general transform relative to the base Gaussian distribution without truncation, scaling or translation parameters—purely the elongation transform effects. In the first set of plots, we consider the effect of the parameter α in the generalized transform. In Figure 5.5, we consider the effect of the parameter β in the generalized transform and in Figure 5.6, we consider the effect of the parameter γ.

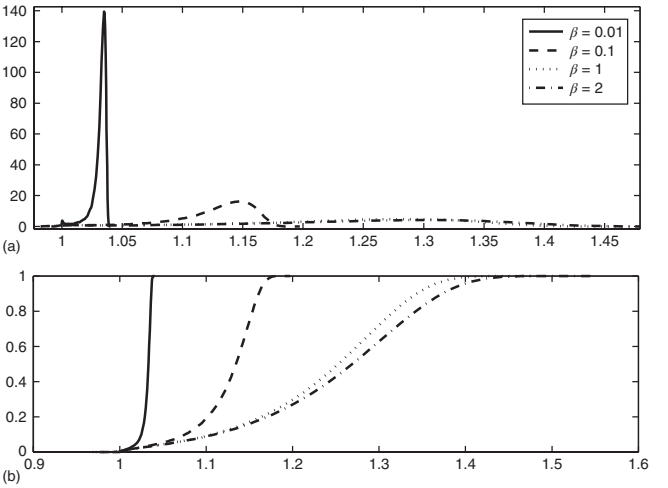

FIGURE 5.5 **(a) The density of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 0.05$, $\beta \in \{0.01, 0.1, 1, 2\}$ and $\gamma = 0.001$. (b) The distribution of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 0.05$, $\beta \in \{0.01, 0.1, 1, 2\}$ and $\gamma = 0.001$.**

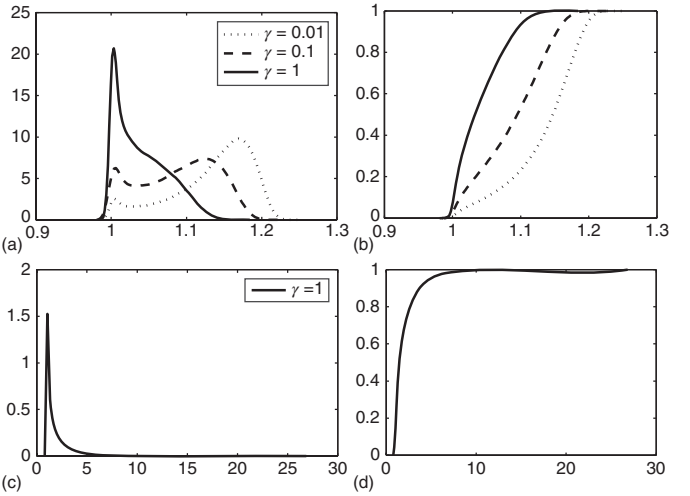

FIGURE 5.6 **(a) The density of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 0.1$, $\beta = 0.1$ and $\gamma \in \{0.01, 0.1, 1\}$. (b) The distribution function of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 0.1$, $\beta = 0.1$ and $\gamma \in \{0.01, 0.1, 1\}$. (c) The density of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 1$, $\beta = 1$ and $\gamma = 1$. (d) The distribution of the T_{hjk} transform with base standard Gaussian distribution and parameters $\alpha = 1$, $\beta = 1$ and $\gamma = 1$.**

These examples simply demonstrate the flexible distributional shapes that can be obtained with the basic elongation transform given by the generalized transform of Fischer (2010) for a different sets of parameter values. In terms of practical severity models, we will now continue to parameterize these transforms to provide sufficient parameters that may make these models suitable for OpRisk loss modeling, leading to, for example, the g-and-h distribution family.

Under the T_{hjk} superclass of transformations, one can state the following basic properties. Assuming that $W \sim Normal(0,1)$ will produce the severity random variable $X = K(W) = WT_{hjk}(W)^{\theta}$, then the severity density and quantile functions for X is given by

$$\begin{aligned} f_X(x) &= \frac{1}{Q'_X(Q_X^{-1}(x))}, \quad \inf\{x : x \in S\} < x < \sup\{x : x \in S\} \\ &= \frac{\phi(K^{-1}(x))}{K'(K^{-1}(x))} \\ Q_X(\alpha) &= K(Q_W(\alpha)), \quad \alpha \in [0,1], \end{aligned} \tag{5.79}$$

where $\phi(\cdot)$ is the standard Normal density function, and S the appropriate support of the random variable X and

$$K'(w) = T_{hjk}(w)^{\theta-1}(T_{hjk}(w) + \theta w T'_{hjk}(w)).$$

Other generalizations to the Tukey family have included those of Rayner & MacGillivray (2002) where they proposed the generalized forms for the quantile functions of the g-and-h and g-and-k families of distributions given in Equations 5.80 and 5.81. The generalized g-and-h and g-and-k families have $b > 0$ and c is the constant to ensure proper distributions are obtained.

$$Q_X(\alpha; a, b, g, h) = a + bQ_W(\alpha)\left(1 + c\frac{1 - \exp(-gQ_W(\alpha))}{1 + \exp(-gQ_W(\alpha))}\right)\exp\left(\frac{1}{2}hQ_W(\alpha)^2\right), \tag{5.80}$$

$$Q_X(\alpha; a, b, g, k) = a + bQ_W(\alpha)\left(1 + c\frac{1 - \exp(-gQ_W(\alpha))}{1 + \exp(-gQ_W(\alpha))}\right)(1 + Q_W(\alpha)^2)^k. \tag{5.81}$$

Next we explain properties of specific sub-families of distributions, showing how these results are derived for the basic g-and-h family.

5.2.1 *g*-AND-*h* SEVERITY MODEL FAMILY IN OpRisk

The family of g-and-h distributions was first studied in Tukey (1977b) and then considered in a number of works such as Hoaglin (2006), Azzalini (1985), Hoaglin (2006) and Fischer *et al.* (2007). The multivariate versions of this model were discussed in Field & Genton (2006).

The advantage of the g-and-h family for modeling severity in an OpRisk LDA framework is due to the fact that it provides a very flexible range of skew, kurtosis and heavy-tailed features whilst also being specified as a rather simple transformation of standard Gaussian random variates, making simulation of the annual loss under such a model efficient and simple. It is important to note that the support of the g-and-h density includes the entire real line, as such one must be cautious in OpRisk settings to manage the treatment of the parameter

settings to restrict the probability of negative values as much as possible. This can be achieved either by truncation or by restriction of the parameter values. In some sub-family members, the g-and-h family is automatically taking a positive support such as the double $h-h$ sub-family. In addition, it has been shown that the g-and-h distribution can approximate most members of the Pearson family of distributions up to a desired level of accuracy.

5.2.1.1 g-and-h, g, h and $h-h$ Family Transformations. The g-and-h family can be considered as composed of three transformations that can produce sub-families of non-Gaussian distributions for severity based on the g-distributions, the h-distributions and the g-and-h distributional families. The basic specifications in which g and h components are treated as constants are given in Definitions 5.20–5.22 in terms of transformations of Gaussian random variables.

Definition 5.20 (g-and-h Distributional Family) *Let $W \sim Normal(0,1)$ be a standard Gaussian random variable. Then the loss random variable X has severity distribution given by the g-and-h distribution with parameters $a, b, g, h \in \mathbb{R}$, denoted by $X \sim GH(a,b,g,h)$, if X is given by (for $g \neq 0$)*

$$X = T_{g,h}(W; a, b, g, h) := a + b\frac{\exp(gW) - 1}{g}\exp\left(\frac{hW^2}{2}\right). \tag{5.82}$$

The parameters a and b are linear transformation parameters and the parameters g and h can be significantly extended to polynomials as discussed later and play an important role in the skewness and kurtosis properties of the g-and-h family.

Remark 5.9 *In general, one may consider the constants g and h above to be more flexibly selected as polynomials which would include higher orders of W^2. These polynomials could take the form, for example, for any integers p and q*

$$\begin{aligned} g(w) &:= \alpha_0 + \alpha_1 w + \cdots + \alpha_p w^p, \\ h(w) &:= \beta_0 + \beta_1 w + \cdots + \beta_q w^q. \end{aligned} \tag{5.83}$$

These polynomial terms can add additional degrees of freedom to improve the ability to fit data. These have been shown to be significant when modeling certain types of OpRisk data, as demonstrated in Dutta & Perry (2006) and Peters & Sisson (2006).

EXAMPLE 5.4 g-and-h Elongation Transform Density Shapes (Base Gaussian Distribution)

Consider the g-and-h elongation transform of a base reference random variable $W \sim Normal(0,1)$ given by

$$X = T_{g,h}(W; a, b, g, h) := a + b\frac{\exp(gW) - 1}{g}\exp\left(\frac{hW^2}{2}\right). \tag{5.84}$$

Figure 5.7 shows the distributions of this general transform relative to the base Gaussian distribution without truncation, scaling or translation parameters—purely the elongation transform effects. In the first set of plots, we consider the effect of the parameter α in the generalized transform.

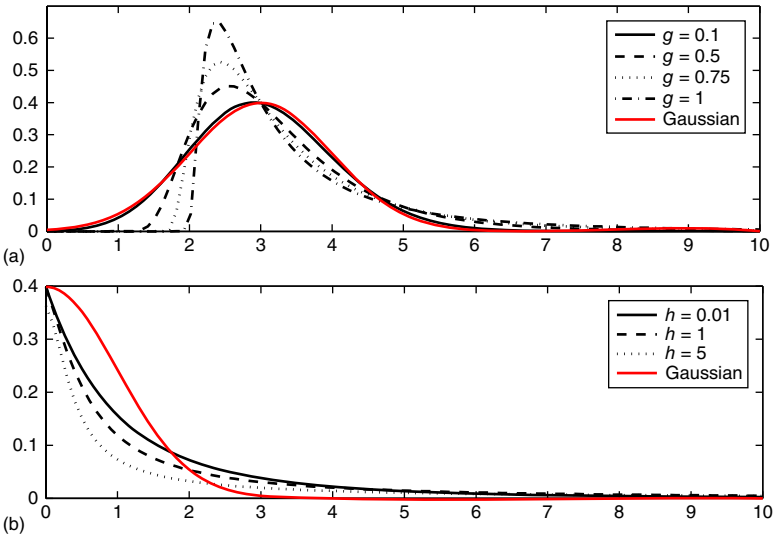

FIGURE 5.7 **(a) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base Gaussian distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (b) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base Gaussian distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$. *See insert for color representation of this figure.***

In this second example, we demonstrate the effect of changing the base distribution to a LogNormal density model instead of the Gaussian distribution.

EXAMPLE 5.5 *g*-and-*h* Elongation Transform Density Shapes: Base Gaussian versus Logistic Distribution

Consider the g-and-h elongation transform of one of the two base reference random variables $W \sim Normal(0, 1)$ or $W \sim LogNormal(0, 1)$ given by

$$X = T_{g,h}(W; a, b, g, h) := a + b\frac{\exp(gW) - 1}{g}\exp\left(\frac{hW^2}{2}\right). \tag{5.85}$$

Figure 5.8 shows the distributions of this general transform relative to the base Gaussian distribution without truncation, scaling or translation parameters—purely the elongation transform effects. In the first set of plots, we consider the effect of the parameter α in the generalized transform.

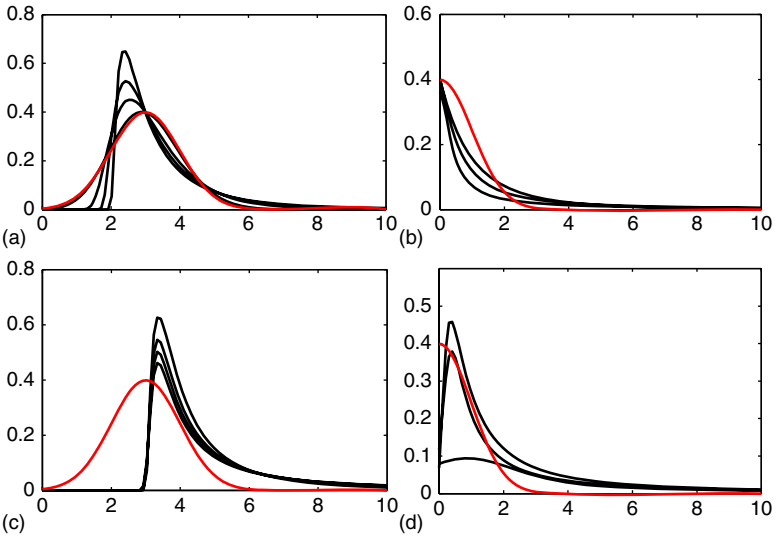

FIGURE 5.8 **(a) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base Gaussian distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (b) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base Gaussian distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$. (c) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base LogNormal(0,1) distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (d) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base LogNormal(0,1) distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$. *See insert for color representation of this figure.***

Remark 5.10 *In Dutta & Perry (2006), it was recommended that a reasonable range of parameter values for the parameters g and h was to restrict them to be $g, h > 0$, and in particular, the enterprise level modeling they performed involved the ranges $g \in [1.79, 2.30]$ and $h \in [0.10, 0.35]$.*

Within this family of g-and-h distributions, one can also define the sub-families of distributions given by the g and the h families. Again, we present these models in their simplest form, with constant g or h, although in practice one may include polynomials in W for such models.

Definition 5.21 (g Distributional Family) *Let $W \sim Normal(0, 1)$ be a standard Gaussian random variable. Then the loss random variable X has severity distribution given by the g-distribution with parameters $a, b, g \in \mathbb{R}$, denoted by $X \sim G(a, b, g)$, if X is given by (for $g \neq 0$)*

$$X = T_g(W; a, b, g) := a + b\frac{\exp(gW) - 1}{g}. \tag{5.86}$$

Remark 5.11 *Note that the g-distribution sub-family corresponds (in the case that g is a constant) to a scaled LogNormal distribution.*

Definition 5.22 (*h* Distributional Family) *Let $W \sim Normal(0,1)$ be a standard Gaussian random variable. Then the loss random variable X has severity distribution given by the h-distribution with parameters $a, b, h \in \mathbb{R}$, denoted by $X \sim H(a,b,h)$, if X is given by*

$$X = T_h(W; a, b, h) := a + bW \exp\left(\frac{hW^2}{2}\right). \tag{5.87}$$

In addition, one may obtain an asymmetric class of $h-h$ distributions studied in Morgenthaler & Tukey (2000, section 2.2), Headrick & Pant (2012a) and Headrick & Pant (2012b). The asymmetric $h-h$ distribution transformation is given in Definition 5.23.

Definition 5.23 (Double *h*–*h* Distributional Family) *Let $W \sim Normal(0,1)$ be a standard Gaussian random variable. Then the loss random variable X has severity distribution given by the unit $h-h$-distribution with parameters $h_l, h_r \in \mathbb{R}$, denoted by $X \sim HH(h_l, h_r)$, if X is given by*

$$X = T_{h,h}(W; h_l, h_r) := \begin{cases} W \exp\left(\frac{1}{2} h_l W^2\right), & W \leq 0, \\ W \exp\left(\frac{1}{2} h_r W^2\right), & W \geq 0, \end{cases} \tag{5.88}$$

for $h_r \geq 0$ and $h_l \geq 0$.

In addition to these families of Tukey transformations discussed, there have been modified g-and-h families developed based on L-moments. The L-moment Tukey transformation families developed in Headrick & Pant (2012b) are based on a transformation of a random variable which involve a logistic distribution, that is, the distribution of random variable W is changed from the standard Gaussian to the following form $W \sim F(w; \mu, s)$ which has density, distribution and quantile functions

$$\begin{aligned} f(w) &= \frac{\exp(-(w-\mu)/s)}{s(1+\exp(-(w-\mu)/s))^2}, \\ F(w) &= \frac{1}{1+\exp(-(w-\mu)/s)}, \\ Q_W(\alpha) &= \mu + s \ln\left(\frac{\alpha}{1-\alpha}\right), \quad \alpha \in [0,1], \end{aligned} \tag{5.89}$$

for all $w \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $s \in \mathbb{R}^+$. The motivation for modifying the distribution transformed under the Tukey structure was related to the fact that inference on the parameters was to be performed with L-moments and L-correlation. The four classes of modified Tukey quantile function transformations are then given in Definition 5.24.

Definition 5.24 (L-Moment Tukey Transforms) *Let $W \sim Logisitc(\mu = 0, s = 1)$ be a standard logistic distributed random variable. Then the loss random variable X has severity distribution given by the L-moment Tukey family as follows:*

1. *The γ–κ Tukey family transformation is given by*

$$X = T_{\gamma,\kappa}(W) = \gamma^{-1}(\exp(\gamma W) - 1)\exp(\kappa|W|). \tag{5.90}$$

This is the analogue of the g-and-h Tukey transform for the logistic distribution case for $\gamma \neq 0$ and $\kappa \geq 0$.

2. *The κ_L–κ_R Tukey family transformation is given by*

$$X = T_{\kappa_L,\kappa_R}(W) = \begin{cases} W\exp(\kappa_L|W|)\,, & W \leq 0, \\ W\exp(\kappa_R|W|), & W \geq 0. \end{cases} \tag{5.91}$$

This is the analogue of the double h–h Tukey transform for the logistic distribution case for $\kappa_L \geq 0$, $\kappa_R \geq 0$ and $\kappa_L \neq \kappa_R$.

Algorithm 5.3 (Simulating Losses from a *g*-and-*h* Severity Model)

1. *Draw a standard Gaussian random variate: $Z_i \sim Normal(0,1)$.*
2. *Given p, q and coefficients $\{\alpha_i\}_{i=0}^{p}$ and $\{\beta_i\}_{i=0}^{q}$, evaluate the polynomials*

$$g(W_i) = \alpha_0 + \alpha_1 W_i + \cdots + \alpha_p W_i^p,$$
$$h(W_i) = \beta_0 + \beta_1 W_i + \cdots + \beta_q W_i^q.$$

3. *Then given parameters a, b and polynomials $g(W_i)$ and $h(W_i)$ evaluate transformation*

$$X_i = a + b\frac{\exp(g(W_i)W_i) - 1}{g(W_i)}\exp\left(\frac{h\,(W_i)\,W_i^2}{2}\right).$$

5.2.1.2 g-and-h, g, h and h–h Family Distribution, Density and Statistical Properties. Next we discuss properties of the g-and-h family and in particular different ways that people have sought to evaluate and present the distribution and density functions for the g-and-h family. In general, it will be informative for this section to remind the reader of the following basic property.

Proposition 5.13 *If X is a continuous random variable distributed according to distribution $X \sim F(x)$ which is monotonically increasing on support $Supp\{F(x)\} = \{x : 0 < F(x) < 1\}$. Then, in this general case, one can show that the quantile function $Q_X(\alpha) = F^{-1}(\alpha)$ for $\alpha \in [0,1]$ determines the relationship between the random variable X and any other continuous random variable with monotonically increasing distribution, say $W \sim G(w)$. The relationship is then specified through the transformation*

$$X \overset{d}{=} F^{-1}(G(W)). \tag{5.92}$$

When the random variable of W is standard Gaussian as utilized in the g-and-h family one can show that for any continuously differentiable transformation $X = T(W)$, then X will have a density given in Equation 5.93 with respect to the standard Gaussian density $\phi(\cdot)$. In this case, one can also observe that when the transform $T(\cdot)$ increases rapidly the resulting density

is heavy tailed. For instance, a linear growth in the function $T(\cdot)$ results in tail behaviour for the distribution of random variable X being equivalent to a Gaussian.

$$f_X(x) = \frac{\phi(T^{-1}(x))}{T'(T^{-1}(x))}. \tag{5.93}$$

As observed previously, the Tukey family has a transformation $T(\cdot)$ given in Definition 5.18, that is,

$$X = T(W) = W\left(\frac{h}{2}W^2\right). \tag{5.94}$$

Note that the original Tukey h-type transformation had $\theta = 1$ and addition scaling of $1/2$ as indicated earlier. This transformation has the property that its derivative

$$\frac{d}{dw}T(w) = (1 + hw^2)\exp\left(\frac{1}{2}hw^2\right) \geq 1 \tag{5.95}$$

for all $h \geq 0$.

In addition, in the following discussions, it will be useful to recall the following properties of the g-and-h family of distributions, see discussions in Dutta & Babbel (2002).

1. The g-and-h transformation can be shown to be strictly monotonically increasing in its argument, that is, for all $w_1 \leq w_2$, one has $T_{gh}(w_1) \leq T_{gh}(w_2)$.
2. If $a = 0$, then the g-and-h transformation satisfies the condition $T_{-g,h}(W) = -T_{g,h}(-W)$.

In Degen *et al.* (2007), it was observed that one can specify the distribution function of the g-and-h distribution as given in Definition 5.25 via a composite function. Note that the scale parameters a and b can be dropped without the loss of generality. The particular representation developed by these authors was convenient to allow on to obtain a closed-form representation of the value-at-risk for such a random variable.

Definition 5.25 (g-and-h Distribution Function (constant g and h with $h > 0$)) *Consider the g-and-h-distributed random variable $X \sim GH(a = 0, b = 1, g, h)$ with constant parameters g and $h > 0$. The distribution function can be specified according to the following composite function:*

$$F_X(x; g, h) = \Phi(k^{-1}(x)), \tag{5.96}$$

where $\Phi(\cdot)$ is the standard Gaussian distribution and the function $k(x)$ is specified by

$$k(x) = \frac{\exp(gx) - 1}{g}\exp\left(\frac{hx^2}{2}\right). \tag{5.97}$$

In this parameterization, the parameter g will control the skew of the distribution in terms of both the sign and the magnitude, whilst the parameter h will control heaviness of the tails and is related directly to the kurtosis. This will be discussed further when the regular variation properties of this model are explored below. Under the restricted parameterization for the distribution given in Definition 5.25, one can obtain the quantile function given in Definition 5.26.

Definition 5.26 (g-and-h Quantile Function (constant g and h with $h > 0$)) *Consider the g-and-h-distributed random variable $X \sim GH(a = 0, b = 1, g, h)$ with constant parameters g and $h > 0$. The quantile function for a level $\alpha \in [0, 1]$ can be specified according to the following representation*

$$Q_X(\alpha) := q_X(\alpha; g, h) = F_X^{-1}(\alpha; g, h) = k(\Phi^{-1}(\alpha)), \tag{5.98}$$

where $\Phi(\cdot)$ is the standard Gaussian distribution and the function $k(x)$ is specified by

$$k(x) = \frac{\exp(gx) - 1}{g} \exp\left(\frac{hx^2}{2}\right). \tag{5.99}$$

Headrick *et al.* (2008) approached the problem of specification of the distribution and density for generic parameterizations of the g-and-h family from an analytic geometry perspective; this can be seen to be an analogous representation of the approach discussed earlier by Degen *et al.* (2007). In fact, the representations they obtained were equivalent with equivalent parameter restrictions. We briefly mention these results later as they allow for an alternative perspective on how one obtains the results in Definitions 5.25 and 5.26. To proceed with the representation developed in Headrick *et al.* (2008), one needs to define $\phi(w) = f_W(w)$ and $\Phi(w) = F_W(w)$, respectively, to be the curves that characterize the standard Gaussian density and distribution. Then considering that w is actually comprised two components (a vector) with an additional auxiliary variable such that $w = (x, y)$ will produce the following mappings of the curves $f_W(w)$ and $F_W(w)$ according to the following relationships:

$$\begin{aligned} f_W(w) : w \to \mathbb{R}^2 &:= f_W(w, f_W(w)), \\ F_W(w) : w \to \mathbb{R}^2 &:= F_W(w, F_W(w)). \end{aligned} \tag{5.100}$$

Given these mappings, one may then define analytically the form of the quantile function for the g-and-h distribution according to the expression given in Equation 5.82 which we reiterate later to present the quantile function notation $Q_X(w)$ of the g-and-h loss random variable given by

$$q(w; g, h) = \frac{\exp(gw) - 1}{g} \exp\left(\frac{hw^2}{2}\right). \tag{5.101}$$

Here, $q(w; g, h)$ is a strictly increasing monotonic function in w with the parameter restrictions $g \neq 0$ and $h > 0$. Using these definitions, Headrick *et al.* (2008) then provide a specification for the density and distribution functions for the g-and-h family as detailed in Definition 5.27.

Definition 5.27 (g-and-h Distribution and Density Functions) *Consider the g-and-h distributed random variable $X \sim GH(a = 0, b = 1, g, h)$ with constant parameters g and $h > 0$. The density and distribution functions associated to the quantile function $q(w; g, h)$ can be specified according to the following composite functions based on the auxiliary variable $w = (x, y)$,*

$$\begin{aligned} f \circ q : q(w; g, h) \to \mathbb{R}^2 &:= f_{q(W;g,h)}(w) = f_{q(W;g,h)}\left(q(w; g, h), \frac{f_W(w)}{q'(w; g, h)}\right), \\ F \circ q : q(w; g, h) \to \mathbb{R}^2 &:= F_{q(W;g,h)}(w) = F_{q(W;g,h)}(q(w; g, h), F_W(w)), \end{aligned}$$

where $q'(z;g,h)$ denotes the derivative of the quantile function given by

$$q'(z;g,h) := \frac{d}{dz}\left[\frac{\exp(gz)-1}{g}\exp\left(\frac{hz^2}{2}\right)\right]$$
$$= \exp\left(gz+\frac{hz^2}{2}\right)+\frac{h}{g}z\exp\left(\frac{hz^2}{2}\right)(\exp(gz)-1).$$

One advantage of the specification of the distribution and density functions with regard to a particular quantile function is that the statistical properties of these distributions can now be easily studied. For instance, the mode and moments of the distribution can be characterized. The result in Proposition 5.14 provides the mode for the g-and-h distribution in Definition 5.27.

Proposition 5.14 (Mode and Median of the g-and-h Density) *Consider the g-and-h-distributed random variable $X \sim GH(a=0,b=1,g,h)$ with constant parameters g and $h>0$. The mode of the density is located at the value $\widetilde{w} = Mode[W]$ which produces a maximum value of the density at $f_{q(W;g,h)}(\widetilde{w})$ and can be found as the solution to the following equation when $w=\widetilde{w}$ which is selected to satisfy*

$$\frac{d}{dw}\left[\frac{f_W(w)}{q'(w;g,h)}\right]=0. \tag{5.102}$$

The median of the g-and-h-distributed random variable is then given by $w_{0.5} = Median[W]$ and will correspond to the median being the limit of $\lim_{w\to 0} q(w;g,h)=0$. Therefore, one sees that in the general g-and-h distribution, the median of the dataset will be the parameter a.

These can be found numerically and for the mode the solution obtained will be guaranteed to be a globally optimal solution, see discussion in Headrick *et al.* (2008). It can also be noted that in the case of the h-type and double h-type Tukey distributions, the median and mode are at the origin (for $a=0$).

In addition, one may now express the moments of the g-and-h-distributed random variable according to the results in Proposition 5.15. It was noted in Dutta & Babbel (2002) that because the g-distribution is a horizontally shifted LogNormal distribution, then the moments of the g-distribution take the same form as those of a LogNormal model with appropriate adjustment for the translation. The h-distributional family is symmetric (except the double $h-h$ family), consequently all odd-order moments for the h-sub-family are zero.

Proposition 5.15 (Moments of the *g*-and-*h* Density) *Consider the g-and-h-distributed random variable $X \sim GH(a=0,b=1,g,h)$ with constant parameters g and $h>0$. The r-th integer moment of the distribution is given with respect to the standard normal distribution and the rth power of the quantile function $q(w)$ by*

$$\mathbb{E}[X^r] = \mathbb{E}[q(W;g,h)^r] = \int_{-\infty}^{\infty} q(w;g,h)^r f_W(w)\,dw, \tag{5.103}$$

which will exist if $h\in[0,1/r)$. One can also observe more generally that under the g-and-h transform, the following identity holds with regard to the powers of the standard Gaussian

$W \sim Normal(0,1)$ *such that*

$$\begin{aligned} X^n &= T_{g,h}(W;a,b,g,h)^n \\ &= (a + bT_{g,h}(W;a=0,b=1,g,h))^n \\ &= \sum_{i=0}^{n} \frac{n!}{(n-i)!i!} a^{n-i} b^i T_{g,h}(W;a=0,b=1,g,h)^i, \end{aligned} \tag{5.104}$$

which will produce moments given by

$$\begin{aligned} \mathbb{E}[X^n] &= \mathbb{E}[(a + bT_{g,h}(W;a=0,b=1,g,h))^n] \\ &= \sum_{i=0}^{n} \frac{n!}{(n-i)!i!} a^{n-i} b^i \mathbb{E}[T_{g,h}(W;a=0,b=1,g,h)^i]. \end{aligned} \tag{5.105}$$

Furthermore, it was shown in Dutta & Babbel (2002) that when it exists, one can obtain the general expression

$$\mathbb{E}[T_{g,h}(W;a=0,b=1,g,h)^i] = \frac{1}{\sqrt{(1-ih)}g^i} \sum\nolimits_{r=0}^{i} (-1)^r \frac{i!}{(i-r)!r!} \exp\left(\frac{(i-r)^2 g^2}{2(1-ih)}\right),$$

which would produce the following four moments in closed form:

$$\begin{aligned} \mathbb{E}[X] = \mathbb{E}[q(W;g,h)] &= \left[\exp\left(\frac{g^2}{2-2h}\right) - 1\right]\left[g\sqrt{1-h}\right]^{-1}, \\ \mathbb{E}\left[X^2\right] = \mathbb{E}\left[q(W;g,h)^2\right] &= \left[1 - 2\exp\left(\frac{g^2}{2-4h}\right)\right. \\ &\quad \left. + \exp\left(\frac{2g^2}{1-2h}\right)\right]\left[g^2\sqrt{1-2h}\right]^{-1}, \\ \mathbb{E}\left[X^3\right] = \mathbb{E}\left[q(W;g,h)^3\right] &= \left[3\exp\left(\frac{g^2}{2-6h}\right) + \exp\left(\frac{9g^2}{2-6h}\right)\right. \\ &\quad \left. - 3\exp\left(\frac{2g^2}{1-3h}\right) - 1\right]\left[g^3\sqrt{1-3h}\right]^{-1}, \\ \mathbb{E}\left[X^4\right] = \mathbb{E}\left[q(W;g,h)^4\right] &= s(g,h)\exp\left(\frac{8g^2}{1-4h}\right)\left[g^4\sqrt{1-4h}\right]^{-1}, \end{aligned}$$

with the function $s(g,h)$ *given by*

$$\begin{aligned} s(g,h) = 1 &+ 6\exp\left(\frac{6g^2}{4h-1}\right) + \exp\left(\frac{8g^2}{4h-1}\right) \\ &- 4\exp\left(\frac{7g^2}{8h-2}\right) - 4\exp\left(\frac{15g^2}{8h-2}\right). \end{aligned}$$

Remark 5.12 *These results allow one to perform model estimation via moment matching of model moments to empirical moments of the loss data.*

As a consequence, one can easily then find the skew, kurtosis and coefficient of variations for the g-and-h distribution as well as the sub-families for the g-distributions and h-distributions. Note that one can also develop variations of the g-and-h distribution density and distribution functions which avoid the restrictions specified in Definitions 5.25 and 5.27. In addition, there has been numerous authors who have studied the generalized properties of quantile-based functionals of asymmetry and kurtosis, see examples in Definition 5.28 and also see Balanda & MacGillivray (1990), Rayner & MacGillivray (2002) and Balanda & MacGillivray (1988).

Definition 5.28 (Generalized Skewness and Kurtosis Functionals in OpRisk) *In considering the generalizations of the skewness and kurtosis for transformation-based quantile function severity models, one can utilize the generalized specifications given for the skewness functional, for a given distribution $F_X(x)$ with respect to its quantile function $Q_X(x)$ by,*

$$\gamma_F = \frac{Q_X(\alpha) + Q_X(1-\alpha) - 2Q_X\left(\frac{1}{2}\right)}{Q_X(\alpha) - Q_X(1-\alpha)}, \quad \alpha \in (0,1). \tag{5.106}$$

In addition, there is the spread functional given by

$$S_F = Q_X(\alpha) - Q_X(1-\alpha), \quad \alpha \in (0,1). \tag{5.107}$$

Such measures were discussed in Balanda & MacGillivray (1990), and it can be shown that $|\gamma_F(\alpha)| \leq 1$. In the case of the g-and-h family of severity models, one would obtain the following forms given in Definition 5.29.

Definition 5.29 (Generalized Skewness and Kurtosis for g-and-h Family) *Consider the g-and-h-distributed random variable $X \sim GH(a = 0, b = 1, g, h)$ with constant parameters g and $h > 0$ and a quantile function for a level $\alpha \in [0,1]$ given by*

$$q_X(\alpha; g, h) = F_X^{-1}(\alpha; g, h) = k(\Phi^{-1}(\alpha)), \tag{5.108}$$

where $\Phi(\cdot)$ is the standard Gaussian distribution and the function $k(x)$ is specified by

$$k(x) = \frac{\exp(gx) - 1}{g} \exp\left(\frac{hx^2}{2}\right). \tag{5.109}$$

Then the generalized skewness and kurtosis are given by

$$\begin{aligned} S_F &= Q_X(\alpha) - Q_X(1-\alpha) \\ &= \frac{\exp\left(g\Phi^{-1}(\alpha)\right) - 1}{g} \exp\left(\frac{h\Phi^{-1}(\alpha)^2}{2}\right) \\ &\quad - \frac{\exp\left(g\Phi^{-1}(1-\alpha)\right) - 1}{g} \exp\left(\frac{h\Phi^{-1}(1-\alpha)^2}{2}\right), \\ \gamma_F &= \frac{Q_X(\alpha) + Q_X(1-\alpha) - 2Q_X\left(\frac{1}{2}\right)}{Q_X(\alpha) - Q_X(1-\alpha)} \end{aligned}$$

$$= \frac{(\exp(g\Phi^{-1}(\alpha)) - 1)\exp\left(\frac{h\Phi^{-1}(\alpha)^2}{2}\right)}{gS_F} + \frac{(\exp(g\Phi^{-1}(1-\alpha)) - 1)\exp\left(\frac{h\Phi^{-1}(1-\alpha)^2}{2}\right)}{gS_F} - 2\frac{(\exp(g\Phi^{-1}(0.5)) - 1)\exp\left(\frac{h\Phi^{-1}(0.5)^2}{2}\right)}{gS_F}.$$

5.2.2 TAIL PROPERTIES OF THE *g*-AND-*h*, *g*, *h* AND *h*-*h* SEVERITY IN OpRisk

In terms of the tail behaviour of the g-and-h family of distributions, the properties of such severity models have been studied by numerous authors such as Morgenthaler & Tukey (2000) and Degen *et al.* (2007). In particular, the tail property (index of regular variation) for the g-and-h family of distributions was first studied for the h-distribution in Morgenthaler & Tukey (2000) and then later for the g-and-h distribution in Degen *et al.* (2007), see Proposition 5.16. In addition, the second-order regular variation properties of the g-and-h family of distributions was studied in Degen *et al.* (2007).

Proposition 5.16 (Index of Regular Variation of *g*-and-*h* Distribution) *Consider the random variable $W \sim Normal(0,1)$ and a loss random variable X which has severity distribution given by the g-and-h distribution with parameters $a, b, g, h \in \mathbb{R}$, denoted by $X \sim GH(a,b,g,h)$, with $h > 0$ and density (distribution) $f(x)$ (and $F(x)$) . Then index of regular variation is given by*

$$\lim_{x\to\infty} \frac{xf(x)}{\overline{F}(x)} = \lim_{x\to\infty} \frac{\phi(u)(\exp(gu) - 1)}{(1-\Phi(u))(g\exp(gu) + hu(\exp(gu) - 1))} = \frac{1}{h} \tag{5.110}$$

for $u = k^{-1}(x)$ where the function $k(x)$ is given by

$$k(x) = \frac{\exp(gx) - 1}{g}\exp\left(\frac{hx^2}{2}\right). \tag{5.111}$$

Hence, one can state that $\overline{F} \in RV_{-1/h}$.

The asymptotic tail behaviour of the h-family of Tukey distributions was studied in Morgenthaler & Tukey (2000, proposition 1) and given in Proposition 5.17.

Proposition 5.17 (*h*-Type Tail Behaviour) *Consider the h-type transformation, where $W \sim Normal(0,1)$ is a standard Gaussian random variable and the loss random variable X has severity distribution given by the h-distribution with parameters $a, b, h \in \mathbb{R}$, denoted by $X \sim H(a,b,h)$, according to*

$$X = T_h(W; a, b, h) := a + bW\exp\left(\frac{hW^2}{2}\right). \tag{5.112}$$

Then the asymptotic tail index of the h-type distribution is then given by $1/h$. *This is equivalent to the g-and-h family for* $g \neq 0$.

This shows that the h-type family has a Pareto heavy-tailed property; hence, the restriction that moments will only exist is of order less than $1/h$. The g-family of distributions can be shown to be sub-exponential in the tail behaviour but not regularly varying. It was shown in (Degen *et al.* 2007, theorem 2.2) that one can obtain an explicit form for the function of slow variation in the g-and-h family as detailed in Theorem 5.3.

Theorem 5.3 (Slow Variation Representation of g-and-h Severity Models) *Consider the random variable* $W \sim Normal(0,1)$ *and a loss random variable* X *which has severity distribution given by the g-and-h distribution with parameters* $a, b, g, h \in \mathbb{R}$, *denoted by* $X \sim GH(a,b,g,h)$, *with* $g > 0$ *and* $h > 0$ *and density (distribution)* $f(x)$ *(and* $F(x)$*)*. *Then* $\overline{F}(x) = x^{-1/h} L(x)$ *for some slowly varying function* $L(x)$ *is given as* $x \to \infty$ *by*

$$L(x) = \frac{h}{\sqrt{2\pi} g^{1/h}} \frac{\left[\exp\left(\frac{g}{h}\sqrt{g^2 + 2h\ln(gx)} - \frac{g^2}{h}\right) - 1\right]^{1/h}}{\sqrt{g^2 + 2h\ln(gx)} - g} \left(1 + O\left(\frac{1}{\ln x}\right)\right).$$

From this explicit Karamata representation developed by Degen *et al.* (2007), they also show that one can obtain the second-order regular variation properties of the g-and-h family.

The implications of these findings are that the g-and-h distribution, under the parameter restrictions $g > 0$ and $h > 0$ belongs to the domain of attraction of an extreme value distribution, such that $X \sim GH(a,b,g,h)$ with distribution F satisfying $F \in MDA(H_\gamma)$ where $\gamma = h > 0$. As a consequence, by the Pickands–Balkema–de Haan theorem, discussed in detail in Cruz *et al.* (2015), one can state that there exists an extreme value index (EVI) constant γ and a positive measurable function $\beta(\cdot)$ such that the following result between the excess distribution of the g-and-h (denoted by $F_u(x) = \Pr[X - u \leq x | X > u]$) and the generalized Pareto distribution (GPD) is satisfied in the tails

$$\lim_{v \uparrow \infty} \sup_{x \in (0,\infty)} |F_u(x) - G_{\gamma,\beta(v)}(x)| = 0. \tag{5.113}$$

For discussion on the rate of convergence in the tails, see Raoult & Worms (2003) and the application of this theorem to the g-and-h case in Degen *et al.* (2007, lemma 3.1) in which it is shown that the order or convergence is given by $O(A\exp(V^{-1}(v)))$ for functions

$$\begin{aligned} V(x) &:= \overline{F}^{-1}(\exp(-x)), \\ A(x) &:= \frac{V''(\ln x)}{V'(\ln x)} - \gamma. \end{aligned} \tag{5.114}$$

Hence, the conclusion from this analysis regarding the tail convergence of the excess distribution of the g-and-h family towards the GPD $G_{\gamma,\beta(v)}(x)$ is given explicitly by

$$\frac{\ln L(x)}{\ln x} \sim \sqrt{2}\frac{g}{h^{\frac{3}{2}}} \frac{1}{\sqrt{\ln(x)}} = O\left(\frac{1}{\sqrt{\ln(k^{-1}(x))}}\right), \quad x \to \infty. \tag{5.115}$$

Remark 5.13 *The implications of this slow rate of convergence are that when data for severities are obtained from a loss process, then if a goodness-of-fit test suggests that one may not reject the null*

hypothesis that this data came from a g-and-h distribution (under a composite test as described in Cruz et al. *(2015)), then one should avoid performing estimation of the extreme quantiles, such as those used to measure the capital via the Value-at-Risk, via methods based on POT's (Peak Over Threshold Models) or EVT based penultimate approximations.*

5.2.3 PARAMETER ESTIMATION FOR THE *g*-AND-*h* SEVERITY IN OpRisk

There have been many proposed methods for performing parameter estimation in the g-and-h family of distributions. In this section, we survey a few of the possibilities. Dutta & Babbel (2002) suggest a method of parameter estimation for constant g and h parameters based on percentile matching. This involves recognizing the following relationships between data percentiles and the g-and-h parameters given in Proposition 5.18.

Proposition 5.18 (*g*-and-*h* Distribution Percentile Matching Estimation) *Consider a g-and-h-distributed random variable $X \sim GH(a, b, g, h)$ and a sample of n loss data points with order statistics $\{X_{(i,n)}\}_{i=1}^{n}$ that will be used to fit the g-and-h distribution. Then the location parameter is given by the following percentile:*

$$\widehat{a} = Median\{X_1, X_2, \ldots, X_n\} = X_{(\lfloor n/2 \rfloor, n)} \tag{5.116}$$

and the value of the g parameter is given by the following estimator (based on the pth percentile given by $p = i/n$ for one or more $i \in \{1, 2, \ldots, n\}$)

$$g_p = -\frac{1}{W_p} \ln\left(\frac{X_{1-p} - X_{0.5}}{X_{0.5} - X_p}\right) \tag{5.117}$$

with $W_p = \inf\{w : \Phi(w) > p\}$ the p-th percentile of a standard normal and $X_p = \inf\{x : F(x) > p\}$ denotes the pth percentile of the sample loss data. It would then be suggested to take a robust estimate of a set of g_p, for a range of percentiles taken from $p \in \{p_L, \ldots, p_U\}$ and then to form the median

$$\widehat{g} = Median\{\widehat{g}_L, \ldots, \widehat{g}_U\}. \tag{5.118}$$

Given $\widehat{g}$, one can estimate the h parameter and the b parameters using the relationship

$$h_p = \frac{2}{W_p^2} \ln\left(\frac{1}{b} \frac{g\left(X_p - X_{1-p}\right)}{\exp(gW_p) - \exp(-gW_p)}\right). \tag{5.119}$$

From this, the estimates of $\widehat{g}$ and $\widehat{h}$ would then allow one to select a range of percentiles taken from $p \in \{p_L, \ldots, p_U\}$ and then regress $Y_p = \ln\left(\frac{g(X_p - X_{1-p})}{\exp(gW_p) - \exp(-gW_p)}\right)$ against the quadrates of the corresponding percentiles of standard Gaussian with scaling given by $W_p^2/2$, the resulting intercept of the regression would be an estimate of $\ln(b)$ and the resulting gradient would be an estimate of h. Given a selection of k percentile levels in $[p_L, p_U]$, the resulting estimators would each be given

by

$$\begin{aligned}\widehat{h} &= \frac{\frac{1}{2}\sum_{p=1}^{k}\left(W_p^2 - \frac{1}{k}\sum_{j=1}^{k}W_j^2\right)\left(Y_p - \frac{1}{k}\sum_{j=1}^{k}Y_k\right)}{\frac{1}{4}\sum_{p=1}^{k}\left(W_p^2 - \frac{1}{k}\sum_{j=1}^{k}W_j^2\right)^2}, \\ \widehat{b} &= \frac{1}{k}\sum_{j=1}^{k}Y_k - \widehat{h}\frac{1}{2k}\sum_{j=1}^{k}W_j^2.\end{aligned} \tag{5.120}$$

In addition, one could perform moment-based matching to estimate the parameters. In this regard, there have been two approaches proposed which utilize the expressions derived earlier for the first four moments which define a system of four non-linear equations that are solved numerically via root search, see Mahbubul *et al.* (2008) for the discussion on the solutions for g and h and then Jiménez & Arunachalam (2011, section 2.2.2) for the simple moment-based estimators for location and scale parameters a and b.

Remark 5.14 *Note that this simple moment-matching-based approach is not particularly recommended as in cases in OpRisk where these models are of major interest; one is typically not only interested in light-tailed models for the severity instead the heavy-tailed high kurtosis severity models are considered. In such cases, it may be that the population moments may not even be finite, although the sample moments will always be finite. The consequence of this is that no matter how much data is utilized to estimate the parameters they will always be biased, or perhaps not well defined under the system of moment conditions.*

The more robust alternative to simple moment matching, especially when the tails of the empirical distribution of the data suggest heavy-tailed features, such as in situations where parameter estimates of h will be large positive values, then it will be numerically more robust to consider an L-moments-based approach such as those proposed in Headrick & Pant (2012b). In the advanced text in Cruz *et al.* (2015), we detail extensively the properties of L-moments estimators for extreme value theory models; we, therefore, defer the reader to this section for details on the estimators we present in Proposition 5.19. Before presenting these results, we briefly recall the definition of the sample L-moments, see Greenwood *et al.* (1979).

Definition 5.30 (Sample L-Moments) *Consider a sample of n observed losses denoted by random variables $\{X_i\}_{i=1}^n$ with associated order statistics in increasing order $\{X_{(i,n)}\}_{i=1}^n$. Then the first four sample L-moments from the data are given by*

$$\begin{aligned}l_1 &= m_0, \\ l_2 &= 2m_1 - m_0, \\ l_3 &= 6m_2 - 6m_1 + m_0, \\ l_4 &= 20m_3 - 30m_2 + 12m_1 - m_0,\end{aligned}$$

with the sample probability moments m_i's given by

$$
\begin{aligned}
m_0 &= \frac{1}{n}\sum_{j=1}^{n} X_{(j,n)}, \\
m_i &= \frac{1}{n}\sum_{j=i+1}^{n} \frac{(j-1)(j-2)\cdots(j-i)}{(n-1)(n-2)\cdots(n-i)} X_{(j,n)}.
\end{aligned}
\tag{5.121}
$$

Next we will define the population L-moments in terms of the parameters of the L-moment γ–κ Tukey family as well as the asymmetric L-moment κ_L–κ_R Tukey family which can be matched to the sample estimated L-moments and then utilized as a system of non-linear equations to be solve numerically via root search for the resulting L-moment parameter estimates.

Proposition 5.19 (L-Moment Estimators for the L-Moment $\gamma - \kappa$ Tukey Family) *Consider a γ-and-κ-distributed random variable $X \sim F(\gamma, \kappa)$ and a sample of n loss data points with order statistics $\{X_{(i,n)}\}_{i=1}^{n}$ that will be used to fit the γ-and-κ distribution. Then under the restrictions that $\gamma + \kappa < 1$, $\kappa < 1$ and $1 + \gamma > \kappa$ which allow the first two L-moments to be finite, one obtains the following two equations for the population first two L-moments λ_1 and λ_2 given by*

$$
\begin{aligned}
\lambda_1 &= \frac{(-\gamma-\kappa)h_1 + (\gamma-\kappa)h_2 + (-\gamma+\kappa)h_3 + 2\kappa h_4 + (\gamma+\kappa)h_5 - 2\kappa h_6}{2\gamma}, \\
\lambda_2 &= \frac{2\gamma - (\gamma+\kappa)^2 h_1 + (\gamma-\kappa)^2(h_1 - h_3) + (\gamma+\kappa)^2 h_5}{2\gamma},
\end{aligned}
$$

where $h_1, h_2, \ldots, h_6$ are defined with respect to the harmonic number functions with the following arguments according to

$$
\begin{aligned}
h_1 &= H\left[\frac{1}{2}(-1-\gamma-\kappa)\right], & h_2 &= H\left[\frac{1}{2}(-1+\gamma-\kappa)\right], \\
h_3 &= H\left[\frac{1}{2}(\gamma-\kappa)\right], & h_4 &= H\left[\frac{1}{2}(-1-\kappa)\right], \\
h_5 &= H\left[-\frac{1}{2}(\gamma+\kappa)\right], & h_6 &= H\left[-\frac{1}{2}\kappa\right],
\end{aligned}
$$

with the harmonic number functions defined for any $x > 0$ by

$$
H[x] := x\sum_{k=1}^{\infty} \frac{1}{k(x+k)}. \tag{5.122}
$$

One can then estimate sample L-moments which can be matched to the population moments to solve numerically for the parameters.

Remark 5.15 *As noted in Headrick & Pant (2012b, p. 9), there are also expressions developed for the population L-skewness τ_3 and L-kurtosis τ_4, should one wish to utilize these for L-moment matching parameter estimation.*

Analogously, the solutions for the first two population L-moments for the class of $\kappa_L - \kappa_R$ Tukey transformations were detailed in Headrick & Pant (2012b) and can be used to perform parameter estimation, as detailed in Proposition 5.20.

Proposition 5.20 (L-Moment Estimators for the L-Moment $\kappa_L - \kappa_R$ Tukey Family) *Consider the asymmetric κ_L-and-κ_R-distributed random variable $X \sim F(\kappa_L, \kappa_R)$ and a sample of n loss data points with order statistics $\{X_{(i,n)}\}_{i=1}^n$ that will be used to fit the κ_L-and-κ_R distribution. Then under the restrictions that $\kappa_L < 1$ and $\kappa_R < 1$ which allow the first two L-moments to be finite, one obtains the following two equations for the population first two L-moments λ_1 and λ_2 given by*

$$\lambda_1 = \frac{1}{4}[2p_5 - 2p_6 - 2p_7 + 2p_8 - \kappa_L p_9 + \kappa_L p_{10} + \kappa_R p_{11} - \kappa_R p_{12}],$$

$$\lambda_2 = \frac{1}{4}[4 + \kappa_L(-4p_5 + 4p_6 + \kappa_L(p_9 - p_{10})) + 4 + \kappa_R(-4p_7 + 4p_8 + \kappa_R(p_{11} - p_{12}))],$$

where $p_5, p_6, \ldots, p_{12}$ are defined with respect to the polygamma functions with the following arguments according to

$$p_5 = P\left[0, \frac{1}{2} - \frac{\kappa_L}{2}\right], \quad p_6 = P\left[0, 1 - \frac{\kappa_L}{2}\right], \quad p_7 = P\left[0, \frac{1}{2} - \frac{\kappa_R}{2}\right],$$

$$p_8 = P\left[0, 1 - \frac{\kappa_R}{2}\right], \quad p_9 = P\left[1, \frac{1}{2} - \frac{\kappa_L}{2}\right], \quad p_{10} = P\left[1, 1 - \frac{\kappa_L}{2}\right],$$

$$p_{11} = P\left[1, \frac{1}{2} - \frac{\kappa_R}{2}\right], \quad p_{12} = P\left[1, 1 - \frac{\kappa_R}{2}\right],$$

with the polygamma functions defined by

$$P[m, x] := (-1)^{m+1} m! \sum_{k=0}^{\infty} \frac{1}{(x+k)^{m+1}}. \tag{5.123}$$

One can then estimate sample L-moments which can be matched to the population L-moments to solve numerically for the parameters.

There have also been approaches based on numerical maximum likelihood applied to the estimation of parameters in the g-and-h family of models, see discussions in Rayner & MacGillivray (2002).

Often in practice, the amount of observed data may not be large; however, there may be reasonable expert opinion available. As such it is often beneficial to adopt a Bayesian estimation framework, as detailed in the next section.

5.2.4 BAYESIAN MODELS FOR THE *g*-AND-*h* SEVERITY IN OpRisk

In this section, we consider two different approaches to constructing Bayesian models for the g-and-h family of severity models. This is particularly important if one wishes to develop an LDA modeling structure that would calibrate such models using a combination of expert opinions as well as collected loss data, as required in Basel II/Basel III standards. The two frameworks we develop can be considered approximations in that the posterior distribution obtained will

be approximate up to any desired level of precision, as specified by the modeller. In particular, we consider first an approach based on the work of Peters & Sisson (2006) which utilizes an approximate Bayesian computation (ABC) formulation, which was the first application of such statistical techniques in finance and risk; then we consider a specially designed conjugate Bayesian formulation based on Askey orthogonal polynomials (see detailed discussion in Cruz *et al.* (2015)).

5.2.4.1 Approximate Bayesian Computation and the g-and-h Severity Model. The basic concept of ABC methods was covered in Cruz *et al.* (2015, chapter 7). Here we briefly review the concept of ABC methods that is growing in popularity in statistics, see Peters & Sisson (2006), Peters *et al.* (2012a), Beaumont *et al.* (2009), Csillery *et al.* (2010), Del Moral *et al.* (2012) and Sisson *et al.* (2010) and the references therein. In particular, we describe the basic Markov chain Monte Carlo (MCMC) sampling methodology first developed in Peters & Sisson (2006) for the g-and-h distribution. In this work, it was recognized that one could exploit the efficiency of simulation from the g-and-h distribution, which only required simulation of a single standard Gaussian random variate for each observation in order to apply ABC methods.

Consider the parameters of the severity model, upon which one wishes to build a Bayesian model, being given in the case of the g-and-h model by vector

$$\boldsymbol{\Theta} = \{a, b, \alpha_0, \alpha_1, \ldots, \alpha_p, \beta_0, \beta_1, \ldots, \beta_q\} \in \Omega,$$

where it is assumed that we condition upon a choice of p and q for the dimension of the $g(z)$ and $h(z)$ polynomials, which can be selected by a desired model selection criterion such as Bayesian information criterion (BIC), Bayes factors or deviance information criterion (DIC). Then, we will denote the prior distribution generically for these severity model parameters by the joint density $\pi(a, b, \alpha_0, \ldots, \alpha_p, \beta_0, \beta_1, \ldots, \beta_q)$. These priors can be elicited in a number of methods *a priori*, see discussions on the different approaches, for instance, in O'Hagan (1998).

Given the prior distribution, essentially, the ABC methods first reduce the observed loss data in a year, denoted by the n losses given by $\boldsymbol{x} = x_{1:n}$, to a low dimensional vector of summary statistics denoted by $t_{\boldsymbol{x}} = T(\boldsymbol{x}) \in \mathcal{T}$, where $\dim(\boldsymbol{\Theta}) \leq \dim(t_{\boldsymbol{x}}) << n$. Then, the true posterior $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is replaced with a new posterior given by $\boldsymbol{\pi}(\theta|t_{\boldsymbol{x}})$ which would theoretically match exactly the true posterior in two cases if $t_{\boldsymbol{x}} = \boldsymbol{x}$ or if $t_{\boldsymbol{x}}$ is sufficient for θ, otherwise it is an approximation $\boldsymbol{\pi}(\theta|t_{\boldsymbol{x}}) \approx \boldsymbol{\pi}(\theta|\boldsymbol{x})$. The new target posterior, still assumed to be computationally intractable (with regard to evaluation of the density pointwise), is embedded within an augmented model from which a Monte Carlo sampling scheme is viable, such as MCMC or sequential Monte Carlo (SMC), see Cruz *et al.* (2015, chapter 7).

The secret to all ABC methods is the replacement of the evaluation of the intractable likelihood model with the simulation of auxiliary data given a set of model parameters $\boldsymbol{\Theta}$. Hence, the auxiliary data will be denoted by vector $\boldsymbol{X}^* = X^*_{1:n}$ and the ith sample is obtained by $X^*_i = x^*_i$ through simulation in the case of the g-and-h model by $X_i \sim GH(a, b, \alpha_0, \alpha_1, \ldots, \alpha_p, \beta_0, \beta_1, \ldots, \beta_q)$ using the algorithm specified in the previous subsection. The auxiliary data is then also summarized by the summary statistic $t_{\boldsymbol{x}^*} \in \mathcal{T}$ for the given simulated realization.

Specifically, under the ABC method, one then expresses the joint posterior of the model parameters $\boldsymbol{\Theta}$ and auxiliary data $\boldsymbol{X}^*$ conditional upon the observed data $\boldsymbol{X}$ according to the

kernel-based representation given by

$$\pi(\theta, \boldsymbol{X}^*|\boldsymbol{x}) \propto K_h(t_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}) f(\boldsymbol{X}|\theta)\pi(\theta), \tag{5.124}$$

where $\boldsymbol{X}^* \sim f(\boldsymbol{X}|\theta)$ and $f(\cdot|\theta)$ is in this case the g-and-h model. Then under this ABC posterior framework, the marginal posterior density is given by

$$\pi_M(\theta|\boldsymbol{x}) = c_M \int K_h(t_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}) f(\boldsymbol{X}|\theta)\pi(\theta) d\boldsymbol{X}^*, \tag{5.125}$$

where

$$c_M^{-1} = \int_\Omega \int K_h(t_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}) f(\boldsymbol{X}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{X}^* d\boldsymbol{\theta}$$

normalizes the posterior such that it is a well-defined density, see discussions in Reeves & Pettitt (2005) and Peters *et al.* (2012d).

The function $K_h(t_{\boldsymbol{x}^*} - t_{\boldsymbol{x}})$ is a standard kernel function with scale parameter $h \geq 0$ which weights the intractable posterior with high density in regions in which $t_{\boldsymbol{x}^*} \approx t_{\boldsymbol{x}}$ where the auxiliary and observed datasets are similar. Therefore, $\pi_M(\boldsymbol{\theta}|\boldsymbol{x}) \approx \pi(\boldsymbol{\theta}|\boldsymbol{x})$ forms an approximation of the intractable posterior through standard smoothing arguments, see Marin *et al.* (2012). As $h \to 0$, so that $K_h(t_{\boldsymbol{x}^*} - t_{\boldsymbol{x}})$ becomes a point mass at the origin (i.e. $t_{\boldsymbol{x}^*} = t_{\boldsymbol{x}}$) and is zero elsewhere, if $t_{\boldsymbol{x}}$ is sufficient for $\boldsymbol{\theta}$, then the intractable posterior marginal is recovered exactly where $\pi_M(\boldsymbol{\theta}|\boldsymbol{x}) = \pi(\boldsymbol{\theta}|\boldsymbol{x})$. Note that, typically, setting h too small is computationally impractical, see discussions in the context of g-and-h models in this regard in Peters & Sisson (2006). There has been a reasonable amount of discussion on the different possible choices one may adopt in practice, see Peters *et al.* (2012d) and discussion therein.

Typically, in practice, it is common to consider a generalization of the above scheme in which the joint posterior distribution in Equation 5.124 is augmented with more than one auxiliary summary vector, by considering for $S \geq 1$ the auxiliary posterior

$$\pi(\boldsymbol{\theta}, \boldsymbol{X}^{*,1:S}|\boldsymbol{x}) \propto \widetilde{K}_h(t^{1:S}_{\boldsymbol{x}^*}, t_{\boldsymbol{x}}) f(\boldsymbol{X}^{1:S}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}), \tag{5.126}$$

where $t^{1:S}_{\boldsymbol{x}^*} = (t^1_{\boldsymbol{x}^*}, t^2_{\boldsymbol{x}^*}, \ldots, t^S_{\boldsymbol{x}^*})$ and for all $i \in \{1, 2, \ldots, S\}$ one has i.i.d. datasets generated from the g-and-h intractable likelihood $\boldsymbol{X}^{*,i} \sim f(\boldsymbol{X}|\boldsymbol{\theta})$. By construction, the auxiliary data are conditionally independent, given $\boldsymbol{\theta}$ which gives

$$f(\boldsymbol{X}^{*,1:S}|\boldsymbol{\theta}) = \prod_{s=1}^{S} f(\boldsymbol{X}^{*,s}|\boldsymbol{\theta}). \tag{5.127}$$

In addition, as discussed in Del Moral *et al.* (2012), one may select the kernel $\widetilde{K}_h(t^{1:S}_{\boldsymbol{x}^*}, t_{\boldsymbol{x}})$ according to

$$\widetilde{K}_h(t^{1:S}_{\boldsymbol{x}^*}, t_{\boldsymbol{x}}) = \frac{1}{S}\sum_{s=1}^{S} K_h(t^s_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}), \tag{5.128}$$

which will result in a joint posterior given by

$$\pi_M(\boldsymbol{\theta}, \boldsymbol{X}^{*,1:S}|\boldsymbol{x}) = c_M \left[\frac{1}{S}\sum_{s=1}^{S} K_h\left(t^s_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}\right)\right] \left[\prod_{s=1}^{S} f\left(\boldsymbol{x}^s|\boldsymbol{\theta}\right)\right] \pi(\boldsymbol{\theta}), \tag{5.129}$$

with the normalizing constant $c_M > 0$. In this case, by construction, one again obtains the appropriate marginal target distribution

$$\int \pi_M \left(\boldsymbol{\theta}, \boldsymbol{X}^{*,1:S} \middle| \boldsymbol{x} \right) \, d\boldsymbol{X}^{*,1:S} = \pi(\boldsymbol{\theta}|\boldsymbol{x}). \tag{5.130}$$

Working with such posterior ABC distributions, one needs to typically obtain samples via a Monte Carlo sampling strategy. In this regard, there are two approaches one may adopt to sample from the target posterior.

1. The first involves treating the summary quantities (statistics) for the auxiliary data, $t_{\boldsymbol{x}^*}^{1:S}$, as parameters in an augmented state space model. This approach, therefore, involves sampling directly on the augmented model, say $\pi_M(\boldsymbol{\theta}, \boldsymbol{X}^{*,1:S}|\boldsymbol{x})$, by obtaining joint samples on the product space $(\boldsymbol{\theta}, \boldsymbol{X}^{*,1:S}) \in \Theta \times \mathcal{X}^S$. Then one *a posteriori* marginalizes over the samples $\boldsymbol{X}^{*,1:S}$ by simply discarding these realizations from the sampler output.
2. The second approach involves sampling the lower dimensional ABC posterior given by $\pi_M(\boldsymbol{\theta}|\boldsymbol{x})$. Within the Monte Carlo sampler, this would involve approximation of the Monte Carlo integral by draws at each iteration of sample for $\boldsymbol{\theta}$, using $\boldsymbol{X}^{*,1}, \boldsymbol{X}^{*,2}, \ldots, \boldsymbol{X}^{*,S} \sim f(\boldsymbol{x}|\boldsymbol{\theta})$ to obtain

$$\begin{aligned} \pi_M(\boldsymbol{\theta}|\boldsymbol{x}) &\propto \pi(\boldsymbol{\theta}) \int_{\mathcal{X}} \widetilde{K}_h(t_{\boldsymbol{x}^*}, t_{\boldsymbol{x}}) f(\boldsymbol{X}|\boldsymbol{\theta}) \, d\boldsymbol{X} \\ &\approx \frac{1}{S} \pi(\boldsymbol{\theta}) \sum_{s=1}^{S} \widetilde{K}_h(t_{\boldsymbol{x}^*}^s, t_{\boldsymbol{x}}) := \widehat{\pi}_M(\boldsymbol{\theta}|\boldsymbol{x}) \end{aligned} \tag{5.131}$$

in lieu of posterior evaluation of $\pi_M(\boldsymbol{\theta}|\boldsymbol{x})$.

To proceed with design of a Monte Carlo sampling strategy to obtain samples from the posterior $\pi_M(\boldsymbol{\theta}|\boldsymbol{x}) \approx \pi(\boldsymbol{\theta}|\boldsymbol{x})$ which is based upon a simple MCMC approach, there are several more advanced strategies available, see, for instance, the SMC samplers—partial rejection control (PRC) method proposed in Peters *et al.* (2012a). A non-complete list of such approaches is provided below:

1. Marginal verses augmented auxiliary ABC posterior: Sisson *et al.* (2010), Sisson & Fan (2011);
2. Rejection, MCMC: Beaumont *et al.* (2002), Marjoram *et al.* (2003);
3. SMC samplers PRC: Peters *et al.* (2012a);
4. SMC samplers: Sisson *et al.* (2007).

The ABC–MCMC algorithm for the g-and-h family then proceeds as follows, given the order of the g and h polynomials p, q.

Algorithm 5.4 (ABC–MCMC for the Bayesian Posterior g-and-h Severity Model)

1. *Initialize the g-and-h model parameters (Markov chain sate):*

$$\boldsymbol{\theta}^{(0)} = \left[\alpha_{1:p}^{(0)}, \beta_{1:p}^{(0)}, a^{(0)}, b^{(0)} \right]$$

and draw synthetic data realizations and evaluate summary statistics

$$t^1_{\boldsymbol{x}^*}, t^2_{\boldsymbol{x}^*}, \ldots, t^S_{\boldsymbol{x}^*} \sim f(t|\boldsymbol{\theta}^{(0)}).$$

This involves for the jth summary vector $t^j_{\boldsymbol{x}^*}$ drawing n samples from the g-and-h model according to

(a) Draw a standard Gaussian random variate: $Z_i \sim Normal(0, 1)$.

(b) Given p, q and coefficients $\{\alpha_i^{(0)}\}_{i=0}^p$ and $\{\beta_i^{(0)}\}_{i=0}^q$, evaluate the polynomials

$$g(W_i) = \alpha_0^{(0)} + \alpha_1^{(0)} W_i + \cdots + \alpha_p^{(0)} W_i^p,$$
$$h(W_i) = \beta_0^{(0)} + \beta_1^{(0)} W_i + \cdots + \beta_q^{(0)} W_i^q.$$

(c) Then given parameters $a^{(0)}$, $b^{(0)}$ and polynomials $g(W_i)$ and $h(W_i)$, evaluate transformation

$$X_i^{(j)} = a^{(0)} + b^{(0)} \frac{\exp(g(W_i)W_i) - 1}{g(W_i)} \exp\left(\frac{h\,(W_i)\,W_i^2}{2}\right).$$

(d) Then given synthetic data samples $\{X_i^{(j)}\}_{i=1}^n$, evaluate the summary statistic $t^j_{\boldsymbol{x}^*}$.

2. *For iterations $k \geq 1$, perform one update on the ABC–MCMC algorithm as follows:*

(a) Generate a proposal vector of parameters for the new state of the Markov chain $\boldsymbol{\theta} \sim q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta})$ using an MCMC proposal for the g-and-h parameters such as a local random walk or a mixture of local and global proposals for $q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta})$.

(b) Draw synthetic data realizations independently from the model with the proposed parameters $\boldsymbol{\theta}$ such that

$$t^1_{\boldsymbol{x}^*}, t^2_{\boldsymbol{x}^*}, \ldots, t^S_{\boldsymbol{x}^*} \sim f(t|\boldsymbol{\theta}).$$

This involves for the jth summary vector $t^j_{\boldsymbol{x}^*}$ drawing n samples from the g-and-h model according to the following steps:

i. Draw a standard Gaussian random variate: $Z_i \sim Normal(0, 1)$.

ii. Given p, q and coefficients $\{\alpha_i\}_{i=0}^p$ and $\{\beta_i\}_{i=0}^q$, evaluate the polynomials

$$g(W_i) = \alpha_0 + \alpha_1 W_i + \cdots + \alpha_p W_i^p,$$
$$h(W_i) = \beta_0 + \beta_1 W_i + \cdots + \beta_q W_i^q.$$

iii. Then given parameters a, b and polynomials $g(W_i)$ and $h(W_i)$, evaluate transformation

$$X_i^{(j)} = a + b \frac{\exp(g(W_i)W_i) - 1}{g(W_i)} \exp\left(\frac{h\,(W_i)\,W_i^2}{2}\right).$$

iv. Then given synthetic data samples $\{X_i^{(j)}\}_{i=1}^n$, evaluate the summary statistic $t^j_{\boldsymbol{x}^*}$.

(c) With probability

$$\min\left\{1, \frac{\frac{1}{S}\sum_s K_h\left(t^s_{\boldsymbol{x}^*} - t_{\boldsymbol{x}}\right)\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(k)})}{\frac{1}{S}\sum_s K_h(t^s_{\boldsymbol{x}^*} - t_{\boldsymbol{x}})\pi(\boldsymbol{\theta}^{(k)})q(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta})}\right\},$$

accept the proposed state and set $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}$ *and keep track of the new sampled values* $\{t^s_{\boldsymbol{x}^*}\}_{s=1}^S$. *Otherwise set* $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)}$.

(d) Increment $k = k + 1$.

Remark 5.16 *The typical summary statistic one may consider involves a summary of the empirical distribution function*

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{x \leq x_i} \tag{5.132}$$

such as a set of sample quantiles $\{\widehat{q}_i\}_{i=1}^p$ *with* $\widehat{q}_i = x_{(i,n)}$. *In addition, one typically has a wide choice of kernel choice such as soft and hard decision kernels see discussions in Peters* et al. *(2010b).*

CHAPTER SIX

Families of Closed-Form Single Risk LDA Models

6.1 Motivating the Consideration of Closed-Form Models in LDA Frameworks

In Cruz *et al.* (2015) and Chapter 4, we have discussed in detail the idea of developing severity models which admit flexible skewness, kurtosis and heavy-tail features for OpRisk modeling. These models represent an important aspect of OpRisk modeling but they comprise only a single component of loss distribution approach (LDA) modeling, and when combined into a compound process setting to obtain the annual loss distribution, several questions naturally arise. The first question to consider involves

Which of these families of models admit a closed-form density and distribution function for the compound process?

In answering this question, one could just enumerate some models with special properties. However, perhaps a more constructive approach would be to consider the question

Under what conditions on the frequency and severity models do closed-form LDA models arise and how best can one characterize such families of OpRisk models?

In characterizing these conditions, one may explain a general LDA model framework that admits closed-form representations for important quantities in OpRisk modeling. The answer to these questions will be addressed in the first part of this chapter.

The second part of this chapter is associated with work with such LDA models in practical settings. It is recognized that even in the case of the single risk LDA model in OpRisk, there will be inherent difficulties associated with calculation numerically of the compound process distribution of an annual loss and the required evaluation of functionals of the tail of the

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

annual loss distribution such as regulatory capital risk measures known as value-at-risk (VaR) and expected shortfall (ES). Several properties of LDA models will first be discussed from a fundamental perspective to provide an understanding of what properties will be required for development of closed-form LDA models. After presenting the theoretical properties, the practical developments will be developed for a range of models so that practitioners may understand in detail the assumptions that must be tested in order for such models to be utilized. Starting with results for univariate settings in which we consider a single risk type and business unit combination. Then we present a few extensions to multivariate results for the multi-risk scenarios.

To summarize briefly, the following topics are covered in this chapter:

1. The characterization of an LDA structure via a convolutional semigroup for doubly infinitely divisible (ID) LDA models.
2. Practical developments for closed-form characterizations of LDA models in both compound and doubly stochastic settings are considered for Poisson, binomial and negative binomial settings.

Each of these models when utilized to construct a family of LDA models will provide analytic exact distributional representations of the compound process LDA models, see discussions in Peters *et al.* (2011). In this class of models, we develop analytic representations for functionals of the annual loss tail quantiles such as the aforementioned risk measures. To complete the chapter, we provide a detailed comparison of all these proposed techniques in a number of practical scenarios.

In this chapter, the primary interest is in obtaining expressions which are exact and closed form for results for single-loss heavy-tailed LDA models. There are a few special cases of OpRisk sub-exponential severity models for which the resulting LDA compound process models are analytic. These are presented later, and then as a consequence of these results and knowledge of the tail approximations or quantile functions, resulting analytic expressions for important risk measures are derived for these models. In particular, Peters *et al.* (2011) present classes of models for which this property is directly relevant to OpRisk modeling. In this section, we consider these models and a few extensions of these models given by

- LDA-Gamma exponential dispersion (ED) family of models which is constructed under a compound process model with frequency from either Poisson or overdispersed Poisson and with severity given by the gamma distributions from Cruz *et al.* (2015);
- LDA-α-stable family of models which is constructed under a compound process model with frequency from Poisson, binomial or negative binomial and with severity given by the α-stable distribution from Chapter 4;
- LDA-tempered-α-stable family of models which is constructed under a compound process model with frequency from Poisson, binomial or negative binomial and with severity given by the tempered-α-stable distribution from Chapter 4;
- LDA-inverse Gaussian (IG) family of models which are constructed under a compound process model with frequency from Poisson, binomial or negative binomial and with severity given by the IG (special member of Tweedie distribution family) or generalized inverse Gaussian (GIG) distribution (special member of generalized hyperbolic family) from Cruz *et al.* (2015).

Each of these models when utilized to construct a family of LDA models will provide analytic exact distributional representations of the compound process LDA models, see discussions in Peters *et al.* 2011. In this class of models, we develop analytic representations for functionals

of the annual loss tail quantiles such as risk measures. To complete the chapter, we provide a detailed comparison of all these proposed techniques in a number of practical scenarios.

6.2 Formal Characterization of Closed-Form LDA Models: Convolutional Semi-Groups and Doubly Infinitely Divisible Processes

OpRisk LDA models are discussed widely in the literature; see, for example, Cruz (2004), Nešlehová *et al.* (2006) Frachot *et al.* (2004) and Shevchenko (2010). Under the LDA Basel II requirements, banks should quantify distributions for frequency and severity of OpRisk for each business line and event type over a 1-year time horizon. In this section, we begin by presenting a generic LDA framework focusing on severity distribution models for which we assume belong to the sub-exponential family of distributions, $F(x) \in \mathscr{F}$.

The standard LDA Basel II/III structures involve an annual loss in a risk cell (business line/event type) modeled as a compound random variable,

$$Z_t^{(j)} = \sum_{s=1}^{N_t^{(j)}} X_s^{(j)}(t). \tag{6.1}$$

Here $t = 1, 2, \ldots, T, T+1$ in our framework is discrete time (in annual units) with $T+1$ corresponding to the next year. The upper script j is used to identify the risk cell. The annual number of events $N_t^{(j)}$ is a random variable distributed according to a frequency counting distribution $P^{(j)}(\cdot)$, typically Poisson. The severities in year t are represented by random variables $X_s^{(j)}(t)$, $s \geq 1$, distributed according to a severity distribution $F^{(j)}(\cdot)$. Severities represent actual loss amounts per event. In addition, in this section, we are particularly interested in situations in which $X_s^{(j)}(t) \sim F^{(j)}(x)$ with the distribution for the severities being a member of the sub-exponential family $F^{(j)}(x) \in \mathscr{F}$, as detailed in Chapter 3.

In proceeding sections, we will further refine this assumption on sub-exponential severity distributions. This will allow us to obtain analytic results which reflect appropriately the nature of OpRisk data that exhibits both extreme but rare events. This is achieved by considering the severity models that exhibit extreme heavy tails, with particular interest in distributions with infinite mean/variance. To accomplish this, we consider the family of α-stable severity distributions. These particular models have been proposed as suitable models for insurance claims modeling and finance, see Embrechts & Schmidli (1994), McNeil & Frey (2000) and Fama & Roll (1968).

For the remainder of this section, our focus will be more generic in nature and we will consider the general assumption that $F^{(j)}(x) \in \mathscr{F}$. We now separate the analysis into two parts, the single risk and multiple risk settings. First, we consider the analysis of the distribution for the annual loss for a given risk process j, that is, we study the distributional properties of random variable $Z_t^{(j)}$ and present appropriate approximations that can be attained for the risk measures based on the annual loss distribution of this compound process random variable. We then consider the distributional properties of multiple risk processes in both an independent setting and a dependent setting. In particular, we present results related to the properties of the distribution of annual loss process from these multiple risks and we provide approximations of relevant risk measures and tail quantiles of the annual loss process in these settings.

We begin the characterization of the family of closed-form LDA models through a theoretically important characterization of such models before proceeding to the implications of

such a representation from a practical perspective. To achieve this, we consider the representation of such models according to their semi-group structure. Therefore, in this section, we draw on the literature that analyses the general operator structures involved in specification of convolutional semi-groups. This theory is developed to explain in a general framework the ability to construct distributions and processes which are closed under convolution, a property that is critical to obtaining closed-form solutions for LDA models. In Feller (1966, chapter IX.2), a detailed analysis of the properties of infinite divisibility of distributions and convolutional semi-groups is discussed. In this section, we precise some of these results to motivate and understand the general features of the forthcoming closed-form models developed.

6.2.1 BASIC PROPERTIES OF CONVOLUTION OPERATORS AND SEMI-GROUPS FOR DISTRIBUTION AND DENSITY FUNCTIONS

We define for a continuous loss random variable the associated severity distribution in $\mathbb{R}$ given by $F(x) = \mathbb{P}\mathrm{r}[X \leq x]$ and density (if the Radon–Nikodym derivative exists) by $f(x)$. We utilize these to define the convolution operation in two settings, the first is between two distributions or densities and the second is between a point function $\phi \in \mathscr{C}[\mathbb{R}]$ in the space of continuous functions $\mathscr{C}$ defined over $\mathbb{R}$ and either a distribution or a density. Note, the following notation for convolutions is utilized specifically in this chapter to distinguish some relevant properties and will return to the standard convolution notation in other chapters.

Definition 6.1 *A convolution between a point function ϕ (either a continuous function or a distribution function) and another distribution is defined to be*

$$T(x) = (F \star \phi)(x) = \int_{\infty}^{\infty} \phi(x-y)\, dF(y), \tag{6.2}$$

where the operator notation $\star$ denotes the presence of a distribution and between a point function ϕ and a density is defined to be

$$t(x) = (f * \phi)(x) = \int_{\infty}^{\infty} \phi(x-y) f(y)\, dy. \tag{6.3}$$

Here, the operator notation $*$ denotes the presence of a density.

From this definition, one may observe the following properties of convolution when ϕ is selected to be a distribution or density, proven in Feller (1966, V.4 theorem 1).

Theorem 6.1 *If ϕ is bounded and continuous, then the resulting function from the convolution $T(x) = (F \star \phi)(x)$ is also bounded and continuous. Furthermore, when ϕ is a probability distribution, then the resulting function $T(x)$ is a probability distribution function.*

This will be of direct consequence to the ability to formulate for an LDA model a closed-form solution firstly to the distribution and density of a compound process (LDA) model for a single risk process. Furthermore, the following properties in Lemmas 6.1 and 6.4 are of relevance to the results to be developed, as discussed in Feller (1966, V.4 theorems 3 and 4).

Lemma 6.1 *When restricted to the space of probability distribution functions, the convolution operator $\star$ is commutative and associative. That is, consider distributions on $\mathbb{R}$ denoted by F_1, F_2 and F_3, then one has:*

- Associative property*: $(F_1 \star F_2) \star F_3 = F_1 \star (F_2 \star F_3)$ for all distributions F_1, F_2 and F_3.*
- Commutative property*: $F_1 \star F_2 = F_2 \star F_1$ for all distributions F_1 and F_2.*

Note that, in general, if ϕ is not a probability distribution function, then the order of operation $(F \star \phi)(x)$ is important. Furthermore, one can state the following additional properties of convolution operators.

Lemma 6.2 *Given two severity distributions on $\mathbb{R}$ denoted by F_1 and F_2, if F_2 is continuous (no atoms), then $T(x) = (F_1 \star F_2)(x)$ is also continuous.*

Lemma 6.3 *Given two severity distributions on $\mathbb{R}$ denoted by F_1 and F_2, if F_2 admits a density $f_2(x)$, then $T(x)$ also admits a density given by $t(x)$.*

Lemma 6.4 *Given two severity distributions on $\mathbb{R}$ denoted by F_1 and F_2 with atoms at points x_1 and x_2, respectively, then $x_1 + x_2$ is an atom of $(F_1 \star F_2)(x)$ and all atoms of $(F_1 \star F_2)(x)$ are characterized in this manner.*

Finally, we also state relevant properties of symmetrization of such an operator when applied to distribution functions, particularly of interest to the sub-family of severity distributions known as α-stable distributions. In (Feller, 1966, V.5 section 5), the following properties of convolution operators are proven with regard to distribution functions (where the notation of Feller (1966) is adopted because of its wide use and clarity).

Lemma 6.5 (Convolutional Symmetrization) *Consider two independent loss random variables X_1 and X_2 with common severity distribution $F_X(x)$. Then the distribution of $X_1 - X_2$ has a symmetric distribution given by ${}^0F(x) = F(x) \star \overline{F}(-x)$.*

Here we utilize the notation adopted in Feller (1966) to denote the symmetrization of $F(x)$ by the distribution ${}^0F(x)$.

Remark 6.1 *This will be of relevance to the family of α-stable models considered below when constructing closed-form compound process LDA models because such severity distributions admit the property that for a standardized $S0$ parameterized density with $\gamma = 1$ and $\delta = 0$, one has*

$$S_\alpha(-x; \beta, 1, 0; S0) = S_\alpha(x; -\beta, 1, 0; S0). \tag{6.4}$$

Clearly, the situation in which a symmetric severity distribution is appropriate will require some discussion in an OpRisk setting, see details below.

Furthermore, the symmetrization property of the convolution operator leads one to a bound on the probability of a given annual loss exceedance for an LDA model given in Theorem 6.2, which can be of direct practical use for an OpRisk practitioner.

Theorem 6.2 (Lower Bounds on the Probability of Exceedance of an Annual Loss) *Given independent loss random variables X_i for $i \in \{1, 2, \ldots, n\}$ from a symmetric severity distribution (typically, in OpRisk a common severity distribution—although not required here), then the probability of the total loss exceeding a quantile q_Z is lower bounded by the following conditions*

$$\begin{aligned} 1 - F_Z(q_Z|N = n) &= \mathbb{P}\mathrm{r}[|X_1 + \cdots + X_n| > q_Z|N = n] \\ &\geq \frac{1}{2}\mathbb{P}\mathrm{r}[\max|X_j| > q_Z|N = n], \end{aligned} \tag{6.5}$$

where if the further restriction of losses being independent and identically distributed (i.i.d.) with $X_i \sim F_X(x)$ applying then one has the analytic bound given by

$$\mathbb{P}\mathrm{r}[\max|X_j| > q_Z|N = n] = 1 - \exp(-n[1 - F(q_Z|N = n) + F(-q_Z|N = n)]). \tag{6.6}$$

This result will be of relevance to construction of a pointwise lower bound on the quantile function of the annual loss distribution as will be explained further in this chapter. Therefore, it will also be of direct interest for estimation of VaR and ES under a closed-form approximation.

Finally, we denote the convolution of a distribution with itself by the notation $F^{(n)*}$ and observe that it satisfies $F^{(n+1)*} = F^{(n)*} \star F$. The analogous notation utilized for an n-fold convolution of densities is then given by $f^{(n)*}$. This is of particular relevance to OpRisk when one considers the addition of loss random variables which are independent. To complete the discussion on fundamental properties of convolution operators, we note that one can define the notion of divisibility of a loss random variables distribution, density or analogously its characteristic function. It is worth noting that we may refer to two concepts here: the first is known as *divisibility* and the second is known as *decomposability* of a random variable, the first being for i.i.d. random variables, whilst the second can distinguish the case of independent but not identically distributed random variables. The two are not exactly the same concept and will have different implications. These have been discussed previously in detail in Chapter 4; we recall a few details again here for convenience. For instance, consider the following two distinguishing features that demonstrate that the condition of divisibility is a stricter requirement to impose than that of decomposability:

- the condition of n-decomposability of a loss random variable X implies that it is also $(n-1)$-decomposable;
- the condition of n-divisibility of a loss random variable X does not necessarily imply that it is $(n-1)$-divisible.

Definition 6.2 (n-Decomposability of Loss Random Variables) *A loss random variable X is said to be n-decomposable if there exists a sequence of independent random variables $\{Y_i\}_{i=1}^n$ such that one can decompose the distribution of the loss random variable according to the following equality in distribution*

$$X \stackrel{d}{=} Y_1 + Y_2 + \cdots + Y_n. \tag{6.7}$$

Definition 6.3 (n-Divisibility of Loss Random Variables) *A loss random variable X is said to be n-divisible if there exists a sequence of i.i.d. random variables $\{Y_i\}_{i=1}^n$ such that one can decompose the distribution of the loss random variable according to the following equality in distribution*

$$X \stackrel{d}{=} Y_1 + Y_2 + \cdots + Y_n. \tag{6.8}$$

It is typically of interest in many applications, especially in OpRisk where the number of losses in the OpRisk reporting period are unknown and random, to consider if there are models

for the severity or the annual loss that will have the special property that they are ID. The class of ID severity distributions in an OpRisk LDA model satisfy the representation of convolution operators as detailed in Definition 6.4, see discussions in Feller (1966, VI.3 p.173).

Definition 6.4 (Infinitely Divisible (ID) Distributions) *A loss random variable X is said to be ID if there exists a sequence of i.i.d. random variables $\{Y_i\}_{i=1}^n$ such that one can decompose the distribution of the loss random variable according to the following equality in distribution*

$$X \stackrel{d}{=} Y_1 + Y_2 + \cdots + Y_n, \forall n \in \mathbb{J}^+. \tag{6.9}$$

Therefore, a severity distribution $X \sim F$ is infinitely divisible if for every integer $n \in \mathbb{J}^+$, there exists a distribution $Y_i \sim F_n$ such that $F = F_n^{(n)*}$ or alternatively one can write the characteristic function of loss random variable X given by $\Phi_X(\theta)$ according to the following power expression of the characteristic function of i.i.d. random variables X_n such that $\Phi_X(\theta) = [\Phi_{X_n}(\theta)]^n$ and the resulting characteristic function will still be a well-defined characteristic function for any $n \in \mathbb{J}^+$.

An important result one can then show is given in Proposition 6.1 which will hold for any infinitely divisible loss random variable, see discussions in Steutel & Van Harn (2003).

Proposition 6.1 (Convolution of Infinitely Divisible Distributions) *Infinite divisibility of loss random variables is preserved under convolution.*

Remark 6.2 *In some cases, the sequence of distributions F_n in the sequence are the same family as F, and in such cases, we say that the random variable is closed under convolution. These classes of models are of direct significance to many applications in OpRisk modeling as will be discussed throughout this chapter.*

The consequence of this for infinitely divisible distributions is that they can always be represented by linear combination of independent random variables with common distribution F_n for any n, see the following example of a gamma-distributed severity.

EXAMPLE 6.1 Infinite Divisibility of a gamma Severity model

Consider a loss random variable $X \sim F$ with the distribution F admitting a density given by a two-parameter (shape, scale) gamma model $X \sim F(x) = Gamma(\alpha, \beta)$ for $\alpha > 0$ and $\beta > 0$. Then the density of the loss random variable is given by

$$f_X(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x) \tag{6.10}$$

and one can show that the n-fold convolution of a gamma density is closed under convolution so the resulting density is also a gamma density with different parameters given by

$$f^{(n)*}(x; \alpha, \beta) = \frac{\beta^{\alpha/n}}{\Gamma\left(\frac{\alpha}{n}\right)} x^{-1+\alpha/n} \exp(-\beta x), \tag{6.11}$$

which is a gamma density with scale parameter β unchanged and shape parameter modified to be $\alpha_n = \alpha/n$. ■

Other severity models that are infinitely divisible and of relevance to OpRisk include

- Poisson distribution;
- Compound Poisson distribution;
- Negative binomial distribution;
- Gamma distribution;
- Pareto distribution;
- LogNormal distribution (Thorin, 1977);
- Half Cauchy distribution;
- Gumbel distribution.

Further more, we can also make general statements about the types of families of distributions that will never be infinitely divisible based on their support, as detailed in Proposition 6.2.

Proposition 6.2 (Non-Existence of Infinite Divisibility on Bounded Support) *A loss random variable that is non-degenerate, that is, it has non-trivial support, but takes a bounded discrete or continuous support cannot be ID.*

In addition, one may further characterize the properties of ID characteristic functions as those that satisfy the result in Proposition 6.3.

Proposition 6.3 (Characteristic Functions of Infinitely Divisible Loss Random Variables) *A loss random variable X which is ID has a characteristic function $\Phi_X(\theta) = \mathbb{E}[\exp(-\mathrm{i}\theta X)]$ which satisfies the property that $\Phi_X(\theta) \neq 0$ for all $\theta \in \mathbb{R}$. In addition, $\Phi_X(\theta)^r$ is a well-defined characteristic function for all $r > 0$.*

Next we state one more useful property of infinite divisibility which we will explore in greater detail throughout sections of this chapter. Namely, if one considers a characteristic function of a compound process, such as one characterizing the annual loss distribution $Z = \sum_{i=1}^{N} X_i$ for i.i.d. losses with frequency distribution given by a Poisson distribution, then the compound process has a characteristic function given by a composite function as follows, where S_n denotes the partial sum $S_n = \sum_{i=1}^{n} X_i$ and one has for the Poisson compound process LDA model the representation

$$\begin{aligned}\Phi_Z(\theta) &= \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N = n]\mathbb{E}[\exp(\mathrm{i}\theta S_n)] \\ &= \mathbb{P}_N(\Phi_X(\theta)) \\ &= \exp(\lambda[\Phi_X(\theta) - 1]),\end{aligned} \tag{6.12}$$

where $\mathbb{P}_N$ is the probability generating function and for $\lambda > 0$ and arbitrary distributions for X. These representations are significant as it can be shown that all infinitely divisible distributions are weak limits of a compound Poisson distribution. To show that this is indeed the case consider a generic representation of a characteristic function for some infinitely divisible random variable X with characteristic function $\Phi_X(\theta)$ and n-th order divisible characteristic function given by

$$\Phi_X(\theta) = [\Phi_{X_n}(\theta)]^n = \exp[n \ln(1 - [1 - \Phi_{X_n}(\theta)])]. \tag{6.13}$$

To see that this will result in the correct specification, one notes that for Φ_{X_n} to be a valid characteristic function, then using the result in Proposition 6.3, one must have the asymptotic behaviour given by $\Phi_{X_n}(\theta) \to 1$ as $n \to \infty$. In addition, one has the asymptotic behaviour $-\ln(1-x) \sim x$ as $x \to 0$, and the conclusion is that one may, therefore, simplify the expression aboves generic representation for n-th order divisibility asymptotically as follows:

$$\Phi_X(\theta) = \lim_{n\to\infty} [\Phi_{X_n}(\theta)]^n = \lim_{n\to\infty} \exp[n(\Phi_{X_n}(\theta) - 1)]. \tag{6.14}$$

This shows that one has obtained in the limit the form of the characteristic function specified for the Poisson compound process. Now in order to take the limit we note that one can rewrite this expression according to the exponential limit of the integral with respect to the distribution for sequence of random variables in the divisibility statement given by distribution $X_n \sim F_n$ as follows:

$$\begin{aligned}
\Phi_X(\theta) &= \lim_{n\to\infty} [\Phi_{X_n}(\theta)]^n \\
&= \exp[n(\Phi_{X_n}(\theta) - 1)] \\
&= \lim_{n\to\infty} \sum_{r=0}^{\infty} \frac{[n(\Phi_{X_n}(\theta) - 1)]^r}{r!} \\
&= \sum_{r=0}^{\infty} \frac{\left[\lim_{n\to\infty} n\left(\Phi_{X_n}(\theta) - 1\right)\right]^r}{r!} \\
&= \exp\left[\lim_{n\to\infty} n\left(\Phi_{X_n}(\theta) - 1\right)\right] \\
&= \exp\left[\lim_{n\to\infty} n \int_{\mathbb{R}} (\exp(\mathrm{i}\theta x) - 1)\, dF_n(x)\right].
\end{aligned} \tag{6.15}$$

To proceed with the representation, it was noted in Steutel & Van Harn (2003) that in general for an arbitrary choice of characteristic function $\Phi_X(\theta)$ when one considers the limit as $n \to \infty$ of nF_n, then it will not converge to a bounded function. However, if the characteristic function of X given by $\Phi_X(\theta)$ has the property that it corresponds to a distribution function that takes a positive mass at $X = 0$ given by p_0 when $n = 0$ and no other atoms are present for the distribution of X, then, in this case, one will have to select the distributions F_n for all n with a jump at $X_n = 0$ given by an amount $p_0^{1/n}$. Then if one defines the standard notation for the characteristic exponentials exponent by

$$\psi_{X_n}(\theta) := \frac{\Phi_{X_n}(\theta) - p_0^{1/n}}{1 - p_0^{1/n}} \tag{6.16}$$

(which is itself a characteristic function) and makes the substitution along with the definition for $\lambda > 0$ that

$$\lambda := -\ln p_0 \tag{6.17}$$

obtained from taking the limit as $n \to \infty$ of $n(1 - p^{1/n})$, then one obtains

$$\begin{aligned}
\Phi_X(\theta) &= \lim_{n\to\infty} [\Phi_{X_n}(\theta)]^n \\
&= \exp\left[\lambda\left(\lim_{n\to\infty} \psi_{X_n}(\theta) - 1\right)\right] \\
&= \exp[\lambda(\psi_X(\theta) - 1)],
\end{aligned} \tag{6.18}$$

where $\psi_X(\theta)$ is the characteristic function of a distribution which does not have mass at the origin and the result is that $\Phi_X(\theta)$ is by definition a compound Poisson characteristic function.

One may now characterize the basic group properties of such convolution operators under certain conditions. Typically, one will be interested in the set of convolution operators applied to point functions or distributions (density) functions on the space $\mathscr{C}(\mathbb{R})$, denoted generically by convolution semi-group operators $\mathscr{D}$. When the identity element $\mathscr{D}(m)\mathbf{1} = \mathbf{1}$ is incorporated into the set, then one naturally can consider a manoid structure to describe the set of such operations. This is a set together with an associative binary operation containing an identity element; note that the manoid may not contain an inverse operation or an identity element and, in such cases, we consider the set of such operators as a semi-group. The remaining section will consider the specific case of the convolutional semi-group (or convolutional associative magma), incorporating the properties of the convolutional operator defined earlier.

Definition 6.5 (Convolution Semi-Group) *For all $m > 0$, consider a probability distribution in $\mathbb{R}$ denoted by $F(x; m)$ and indexed by m. Furthermore, define the operator associated with all such probability distributions, defined for a point function $\phi \in \mathscr{C}(\mathbb{R})$, by*

$$\mathscr{D}(m; F)\phi(x) = \int_{-\infty}^{\infty} \phi(x - y)\, dF(y; m), \tag{6.19}$$

such that the convolution between two severity distributions $F(x; m) \star F(x; n) = F(x; m + n)$ is equivalent in operator to $\mathscr{D}(m; F)\mathscr{D}(n; F) = \mathscr{D}(m + n; F)$. Then the family (set) of all such operators $\{\mathscr{D}(m; F) : \ m \geq 0\}$ is a convolution semi-group associated with the distribution F.

To illustrate this concept, consider Example 6.2 which presents the convolution semi-group for the family of ID distributions comprised the Normal-inverse-Gaussian (NIG) severity models. Such models hold the interesting distinction of being the only members of the hyperbolic distributions to be closed under convolution and hence of direct interest in this chapter for closed-form LDA models.

EXAMPLE 6.2 Normal-Inverse-Gaussian Convolution Semi-Group

An example of such a convolution semi-group of relevance to OpRisk is the family of NIG models. Consider a loss random variable $X \sim NIG(x; \alpha, \beta, \delta, \mu)$ with tail exponent (steepness) α, symmetry β, scale δ and location μ, then the density is given by

$$f_X(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta \mathrm{K}_1(\alpha\sqrt{\delta^2 + (x - \mu)^2})}{\pi\sqrt{\delta^2 + (x - \mu)^2}} \exp(\delta\sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)), \tag{6.20}$$

with $x \in \mathbb{R}$, $\alpha > 0$, $\delta > 0$, $\mu \in \mathbb{R}$ and $0 < |\beta| < \alpha$ and $\mathrm{K}_1(x)$ the modified Bessel function of the third kind with index 1. This family of severity distributions satisfy the following properties of scaling and convolution, see Eriksson *et al.* (2009). If $X \sim NIG(\alpha, \beta, \delta, \mu)$, then the scaled version $cX \sim NIG(\alpha/c, \beta/c, c\delta, c\mu)$ and, furthermore, the convolution of the distributions satisfy

$$(NIG(\alpha, \beta, \delta_1, \mu_1) \star NIG(\alpha, \beta, \delta_2, \mu_2))(x) = NIG(x; \alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2).$$

Hence, considering the family of distributions $F(x;\delta_n,\mu_n) = NIG(x;\alpha,\beta,\delta_n,\mu_n)$ for the continuous range of admissible parameters δ_n,μ_n will produce a convolution semi-group for operators of the form

$$\mathscr{D}(m;F)\phi(x) = \int_{-\infty}^{\infty} \phi(x-y)\,dF(y;\delta_m,\mu_m),$$

for a point function $\phi \in \mathscr{C}(\mathbb{R})$ such that $\mathscr{D}(m;F)\mathscr{D}(n;F) = \mathscr{D}(m+n;F)$ as required. ■

For a second illustration based on the IG distribution, see Example 6.3. Note this member has a strictly positive support and so is suitable for OpRisk severity modeling without any restrictions; however, such a family of distributions does not possess the heavy-tailed properties of the extreme consequence results discussed primarily in this chapter. This distribution does, however, enjoy an extensive history in the statistical literature, it has been named the inverse Gaussian (IG) family of distributions by Tweedie in 1945 and forms a special member of the Tweedie distributions a sub-family of the exponential family of distributions when the index parameter is $p = 3$, see Tweedie (1984); analogously, it is also considered as a special case of the Halphen distributional family, see discussions in Perreault *et al.* (1999).

EXAMPLE 6.3 Inverse-Gaussian Convolution Semi-Group

A second example of such a convolution semi-group of relevance to OpRisk in the family of IG severity models (although this is not a sub-exponential model, it will still be suitable for many LDA single risk processes). There are many parameterizations of the Tweedie distribution, see a discussion in an insurance context in Peters *et al.* (2009b). However, in general, one can consider a Tweedie severity density as given by

$$f_X(x;\mu,\phi,p) = a(x,\phi)\exp\left(\frac{1}{\phi}\left[x\theta(\mu) - \kappa\left(\theta(\mu)\right)\right]\right) \tag{6.21}$$

with

$$\theta(\mu) = \begin{cases} \dfrac{\mu^{1-p}-1}{1-p}, & p \neq 1, \\ \ln\mu, & p = 1, \end{cases}$$

$$\kappa(\theta(\mu)) = \begin{cases} \dfrac{\mu^{2-p}-1}{2-p}, & p \neq 2, \\ \ln\mu, & p = 2, \end{cases}$$

where the special closed-form density case of $p = 3$ produces the IG severity density and distributions reparameterized with mean parameter $\mu > 0$ and shape parameter $\lambda > 0$ according to

$$f_X(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), x \geq 0,$$

$$F_X(x) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right), \quad (6.22)$$

where $\Phi(\cdot)$ is the standard Gaussian distribution. This family of severity distributions satisfy the following properties of scaling and convolution, see the review in Folks & Chhikara (1978). If $X \sim IG(\mu, \lambda)$, then the scaled version $cX \sim IG(c\mu, c\lambda)$ and, furthermore, the convolution of the distributions satisfies (for common shape and scale) according to

$$(IG(\mu_1, \lambda_1) \star IG(\mu_1, \lambda_1))(x) = IG(x; 2\mu_1, 2\lambda_1). \quad (6.23)$$

Hence, considering the family of distributions $F(x; \mu_n, \lambda_n) = IG(x; \mu_n, \lambda_n)$ for the continuous range of admissible parameters μ_n, λ_n will produce a convolution semi-group for operators of the form

$$\mathscr{D}(m; F)\phi(x) = \int_{-\infty}^{\infty} \phi(x - y)dF(y; \delta_m, \mu_m), \quad (6.24)$$

for a point function $\phi \in \mathscr{C}(\mathbb{R})$ such that $\mathscr{D}(m; F)\mathscr{D}(n; F) = \mathscr{D}(m + n; F)$ as required. ■

It is then apparent that such operators are typically termed transition operators, because it is clear that for a function $\phi(x)$ that satisfies $0 \leq \phi(x) \leq 1$ for all $x \in \mathbb{R}$, then $0 \leq \mathscr{D}(m; F)\phi(x) \leq 1$, and in addition, the operator applied to the identity returns the identity $\mathscr{D}(m; F)\mathbf{1} = \mathbf{1}$. This is clearly of direct importance to the consideration of such operators applied to the setting in which the point function ϕ is selected to be a severity distribution function.

One can also characterize uniquely the convolution semi-group for a particular family of severity distribution models in an LDA framework through specification of the generator of the semi-group, as detailed in (Feller 1966, IX.2 definition 4 p.285) and given in Definition 6.6.

Definition 6.6 (Generator of Convolution Semi-Group) *A convolution semi-group* $\{\mathscr{D}(m; F) : m \geq 0\}$ *for a severity distribution* F *is said to be generated by an operator* $\mathscr{G} : \mathscr{C}^{\infty}(\mathbb{R}) \to \mathscr{C}^{\infty}(\mathbb{R})$ *if the following limit is satisfied by the operators:*

$$\lim_{m \downarrow 0} \frac{\mathscr{D}(m; F) - \mathbf{1}}{m} = \mathscr{G}(F), \quad (6.25)$$

where $\mathbf{1}$ *denotes the identity function.*

A consequence of this is that a semi-group with a generator is continuous, where a continuous semi-group is defined as in Definition 6.7.

Definition 6.7 (Continuous Convolution Semi-group) *A convolution semi-group* $\{\mathscr{D}(m; F) : m \geq 0\}$ *for a severity distribution* F *is said to be continuous if the following limit applies:*

$$\lim_{m \downarrow 0} \mathscr{D}(m; F) = \mathbf{1}, \quad (6.26)$$

which implies that $\mathscr{D}(0; F) = \mathbf{1}$.

In the following examples, we provide in the context of OpRisk several classes of important LDA models that will admit interesting and directly relevant convolution semi-groups leading to closed-form expressions for many important quantities in OpRisk quantification for single risk processes. In particular, we specify such semi-groups through the existence of their generator.

EXAMPLE 6.4 Convolutional Semi-group Generator for Truncated Poisson Process LDA

Consider an OpRisk LDA model with a homogeneous compound Poisson process model. In this case, frequency distribution for the number of events in a given interval $[t, t+\tau]$ is assumed to be modeled by a truncated Poisson process such that $[N_{t+\tau} - N_t] \sim Poission\,(N; \lambda\tau, m_\tau)\mathbb{I}[N_{t+\tau} - N_t \leq m_\tau]$ and the truncation level can be some function of the time increment as denoted by $m_\tau \in \mathbb{J}^+$. Then the loss on any given event in the period $[t, t+\tau]$ is modeled by a severity distribution constant over time in which each loss is i.i.d. with $X_i \sim F_X(x)$, then the total loss described by the truncated compound Poisson process distribution is given by

$$\begin{aligned} Z_{[t,t+\tau]} &= \sum_{i=1}^{N_{[t,t+\tau]}} X_i \sim G(z; m_\tau, \tau) \\ &= \sum_{n=0}^{m_\tau} \mathbb{P}\mathrm{r}[N_{[t,t+\tau]} = n | N_{[t,t+\tau]} < m_\tau] F^{(n)\star}(z) \\ &= \left(\sum_{s=0}^{m_\tau} \frac{(\lambda\tau)^s}{s!} \right)^{-1} \sum_{n=0}^{m_\tau} \frac{(\lambda\tau)^n}{n!} F^{(n)\star}(z) \\ &= Z(m_\tau, \lambda\tau)^{-1} \sum_{n=0}^{m_\tau} \frac{(\lambda\tau)^n}{n!} F^{(n)\star}(z). \end{aligned} \tag{6.27}$$

Then it can be shown that for any point function $\phi \in \mathscr{C}(\mathbb{R})$ on the total loss in time interval $[t, t+\tau]$, the resulting semi-group characterized by the integral operation of

$$\mathscr{D}(\tau; G)\phi(z) = \int_{-\infty}^{\infty} \phi(z-y)\, dG(y; m_\tau, \tau) \tag{6.28}$$

is specified by a generator which is obtained by considering the Definition 6.6 for the convolution operator with respect to the distributional family G. To find this generator, we consider

$$\begin{aligned} &\mathscr{D}(\tau; G)\phi - \phi = (Z(m_\tau, \lambda\tau)^{-1} - 1)\phi + Z(m_\tau, \lambda\tau)^{-1}\lambda\tau \\ &\times \left[(F \star \phi)(z) + \frac{\lambda\tau}{2!}(F^{(2)\star} \star \phi)(z) + \cdots + \frac{(\lambda\tau)^{m_\tau - 1}}{m_\tau!}(F^{(m_\tau)\star} \star \phi)(z) \right] \end{aligned} \tag{6.29}$$

and then by dividing by τ and taking the limit as $\tau \downarrow 0$ to obtain the generator,

$$\lim_{\tau \downarrow 0} \frac{\mathscr{D}(\tau; G)\phi - \phi}{\tau} = \lim_{\tau \downarrow 0} \left\{ \frac{\left(Z(m_\tau, \lambda\tau)^{-1} - 1\right)\phi}{\tau} + Z(m_\tau, \lambda\tau)^{-1}\lambda \right.$$

$$\times \left[(F \star \phi)(z) + \frac{\lambda\tau}{2!}(F^{(2)\star} \star \phi)(z) + \cdots + \frac{(\lambda\tau)^{m_\tau - 1}}{m_\tau!}(F^{(m_\tau)\star} \star \phi)(z) \right] \Bigg\}$$

$$= \lim_{\tau \downarrow 0} \left\{ \frac{\left(Z(m_\tau, \lambda\tau)^{-1} - 1\right)\phi}{\tau} + Z(m_\tau, \lambda\tau)^{-1}\lambda(F \star \phi)(z) \right\}, \tag{6.30}$$

where one observes that because $\lim_{\tau \downarrow 0} \mathbb{P}\mathrm{r}[N_{[t,t+\tau]} = 0 | m_\tau] \to 1$ for all $m_\tau \in \mathbb{J}^+$ and furthermore that $\lim_{\tau \downarrow 0} Z(m_\tau, \lambda\tau)^{-1} \to 1$ for all $m_\tau \in \mathbb{J}^+$. Hence, if one considers the point function ϕ given by the identity $\mathbf{1}$, then one has according to Definition 6.6 the generator for this LDA models semi-group given by the operator

$$\begin{aligned} \mathscr{G}(G)(z) &= \lim_{\tau \downarrow 0} \left\{ \frac{\left(Z(m_\tau, \lambda\tau)^{-1} - 1\right)\mathbf{1}}{\tau} + Z(m_\tau, \lambda\tau)^{-1}\lambda(F \star \mathbf{1})(z) \right\} \\ &= \lambda((F \star \mathbf{1})(z) - \mathbf{1}), \end{aligned} \tag{6.31}$$

which gives the generator according to the operator $\mathscr{G} = \lambda(\mathscr{F} - \mathbf{1})$, where $\mathscr{F}$ denotes a convolution operator with respect to the severity distribution F. Note that this has an identical form to the generator for the untruncated Poisson process case. ■

In general, one can also state the following properties of translations and additions of generator operators and the resulting semi-groups they generate under such situations as detailed in Lemma 6.6, see Feller (1966, IX.2 p.286 Volume 2).

Lemma 6.6 *Given two generators $\mathscr{G}_1$ and $\mathscr{G}_2$ for convolution semi-groups $\{\mathscr{D}_1(m, F)\}$ and $\{\mathscr{D}_2(m, F)\}$, respectively, the addition of $\mathscr{G}_1 + \mathscr{G}_2$ also generates a convolution semi-group of operators $\mathscr{D}(m, F) = \mathscr{D}_1(m, F)\mathscr{D}_2(m, F)$.*

Lemma 6.7 *Given a generator $\mathscr{G}$ for a convolution semi-group $\{\mathscr{D}(m, F)\}$, then the translation of the generator by $\mathscr{G} - \alpha\frac{d}{dx}$ generates a new semi-group $\{\widetilde{\mathscr{D}}(m, F)\}$ such that the family of distributions associated with convolution semi-group $\{\mathscr{D}(m, F)\}$, denoted by $F(x; m)$, are translated versions of those distributions representing semi-group $\{\widetilde{\mathscr{D}}(m, F)\}$, given by $\widetilde{F}(x;\ m) = F(x - \alpha m;\ m)$.*

Many other properties are known for the characterization of such semi-groups and their representative generators. For example, distinct convolution semi-groups must have a unique generator and special representations are obtained for convolution operators and their semi-groups, considered with respect to distributions with finite variance. However, such finite variance settings are not necessarily of interest in the context of this chapter, see discussion in Feller (1966, IX.4 section 4 p.289).

From these properties, one can state the following two key results for convolutional semi-groups. In Theorem 6.3, a generic representation of the generator of any continuous semi-group is provided. Then a set of equivalence results are provided for understanding the relationships between probability distributions in the LDA model structure and their representations.

Theorem 6.3 (Generic Convolution Semi-Group Generator) *Consider a distribution function $F(x; m)$ indexed by m for which a continuous convolution semi-group $\{\mathscr{D}(m, F)\}$*

is defined with respect too. If for this distribution $F(x;m)$ *one defines the distribution* $\widetilde{F}(x;m) = \frac{x^2}{m}F(x;m)$ *as* $m \to 0$ *such that it satisfies the condition*

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2} d\widetilde{F}(x;m) < \infty, \tag{6.32}$$

then any continuous convolution semi-group given with respect to this distribution F *by* $\{\mathscr{D}(m,F)\}$ *will have a unique generator* $\mathscr{G}$ *of the form given by*

$$\mathscr{G} = \mathscr{G}^{(0)} - \alpha \frac{d}{dx}, \tag{6.33}$$

where the base operator $\mathscr{G}^{(0)}$ *is defined with respect to a point function* $\psi \in \mathscr{C}^{\infty}(\mathbb{R})$ *such that*

$$\mathscr{G}^{(0)}\phi(x) = \int_{-\infty}^{\infty} \frac{\phi(x-y) - \phi(x) + \frac{d}{dx}\phi(x)T_s(y)}{y^2} d\widetilde{F}(y), \tag{6.34}$$

with truncation function

$$T_s(x) = \begin{cases} s, & x \geq s, \\ x, & |x| \leq s, \\ -s, & x \leq s, \end{cases} \tag{6.35}$$

and a constant α.

A consequence of this result is the following characterization of infinitely divisible distributions given in Theorem 2.4, see Feller (1966, IX.5, p.294). There is also a more modern treatment and detailed discussion of these points found in Cont & Tankov (2004, chapter 2) and Sato (1999). This representation was also discussed in the derivation of the α-stable and tempered stable (TS) models in Chapter 4. Here we treat such a representation in a more general manner.

Theorem 6.4 (Equivalencies of Infinitely Divisible Distributions) *Equivalence between the following representations of ID distributions holds:*

1. *Distributions* $F(x;m)$ *associated with continuous convolution semi-groups with operators* $\mathscr{D}(mF)$*;*
2. *Distributions* $F(x;m)$ *of increments of stochastic processes with stationary-independent increments;*
3. *Limits of sequences of compound Poisson distributions;*
4. *ID distributions and limits of such distributions.*

It is this equivalence that allows one to consider characterizing an important set of closed-form Poisson compound process models for LDA modeling according to a Lévy process representation as follows. Consider the basic characterization of a Lévy process given in Definition 2.6, see Cont & Tankov (2004, definition 3.1). In general, in this section, the primary interest is in processes in $\mathbb{R}$, although all results presented here are extendable to $\mathbb{R}^d$ for any $d \in \mathbb{J}^+$.

Definition 6.8 *A stochastic process which is continuous from the right and has limits existing from the left (*càdlàg*) denoted by* $\{X_t\}_{t\geq 0}$ *with* $X_t \in \mathbb{R}$ *for all* $t \in [0,\infty)$*, defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a Lévy process if the following properties are satisfied:*

- **Independent Increments.** *If* $X_0 = 0$ *and increments of the process* X_0, $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ *are independent.*
- **Stationarity of Increments.** *If the law of increments* $X_{t+\tau} - X_t$ *is independent of* t *and only depends on the time difference* τ*.*
- **Stochastic Continuity.** *If for all* $\varepsilon > 0$*, the following limit exists*

$$\lim_{\tau \to 0} \mathbb{P}(|X_{t+\tau} - X_t| \geq \varepsilon) = 0. \tag{6.36}$$

Furthermore, such a stochastic process will always have a characteristic function given by

$$\mathbb{E}[\exp(\mathrm{i}\theta X_t)] = \exp(t\psi(\theta)),\ u \in \mathbb{R}, \tag{6.37}$$

for a function known as the characteristic exponent of X *which is the map* $\psi : \mathbb{R} \mapsto \mathbb{R}$*.*

If we define the Lévy measure according to Definition 6.9 which measures the average number of jumps in a unit interval of time within a Borel set A from the increments of a Lévy process. Such a measure can then be utilized to re-express the characteristic function of the Lévy process, therefore, because of equivalence, the characteristic function of all infinitely divisible distributions and, therefore, members of the family of distributions $F(x;m)$ associated with continuous convolution semi-groups with operators $\mathscr{D}(mF)$. This is achieved by obtaining a representation of the characteristic exponent function ψ (Lévy-Khinchin representation, see chapter 4) in terms of three defining components known as the Lévy triplet (σ, γ, W) in Theorem 2.5, see Cont & Tankov (2004) and Sato (1999).

Definition 6.9 (Lévy Measure) *If* $\{X_t\}_{t\geq 0}$ *is a Lévy process on* $\mathbb{R}$*, then the measure on* $\mathbb{R}$ *defined by the expected number, per unit time, of jumps whose size belongs to a Borel set* $B \in \mathscr{B}(\mathbb{R})$*, is given by*

$$W(A) = \mathbb{E}\left[\sum_{i=0}^{N_{[0,1]}} \mathbb{I}\left[\left\{X_{\tau_i} - X_{\tau_{i-1}} \in B\right\} \text{ and } \left\{X_{\tau_i} - X_{\tau_{i-1}} \neq 0\right\}\right]\right], \tag{6.38}$$

where the Lévy process has jumps at times $\{\tau_i\}_{i=1:N_t}$ *in time increment* $[0,t]$*.*

Furthermore, several characterizations of the Lévy process have been derived as a consequence of the equivalencies presented in Theorem 6.4. The following Theorem 6.5 presents the Lévy–Khinchin representation of a Lévy process in terms of the Lévy measure. Then in Theorem 6.5, the Lévy-Ito decomposition of a Lévy process is presented.

Theorem 6.5 (Lévy–Khinchin Characteristic Function for Lévy Process) *Given a Lévy process* $\{X_t\}_{t\geq 0}$ *on* $\mathbb{R}$*, then the Lévy-Khinchin representation of the characteristic function is given by*

$$\mathbb{E}[\exp(\mathrm{i}\theta X_t)] = \exp(t\psi(\theta)),\ \theta \in \mathbb{R}, \tag{6.39}$$

with

$$\psi(\theta) = -\frac{1}{2}\sigma^2\theta^2 + \mathrm{i}\gamma\theta + \int_{\mathbb{R}}(\exp(\mathrm{i}\theta x) - 1 - \mathrm{i}\theta x\mathbb{I}[|x| \leq 1])W(dx), \tag{6.40}$$

where $\sigma^2 \in \mathbb{R}^+$, $\gamma \in \mathbb{R}$ *and the Lévy measure* $W(dx)$ *of ID distribution* F *is a positive measure satisfying*

$$\int_{|x|\leq 1}|x|^2 W(dx) < \infty, \quad \int_{|x|\geq 1} W(dx) < \infty. \tag{6.41}$$

Next we note that the infinitesimal generator of a Lévy process, which corresponds to the generator of the convolutional semi-group discussed at the start of this chapter, is given in Proposition 6.4, see discussion in Cont & Tankov (2004, proposition 3.16).

Proposition 6.4 (Generic Convolutional Semi-group Generator of Lévy Process) *Given a Lévy process* $\{X_t\}_{t\geq 0}$ *with triplet* $(\sigma, \gamma, W(dx))$*, the infinitesimal generator of the convolutional semi-group characterizing the Lévy process is given for any point function* $\phi \in \mathscr{C}_0^2(\mathbb{R})$ *with respect to the Lévy triplet by*

$$\begin{aligned}\mathscr{G}\phi(x) = &\frac{1}{2}\sigma^2\frac{d^2\phi(x)}{dx^2} + \gamma\frac{d\phi(x)}{dx} \\ &+ \int_{-\infty}^{\infty}\left(\phi(x+y) - \phi(x) - y\frac{d\phi(x)}{dx}\mathbb{I}[|y| \leq 1]\right)W(dy),\end{aligned} \tag{6.42}$$

and it is observed that the Lévy measure $W(dy)$ *plays the role of* $d\widetilde{F}(y)$ *in the generic specification in Theorem 6.3.*

In addition, one can state the following properties on the moments of the ID random variables that have the Lévy–Khinchin characteristic function representation as given in Proposition 6.5.

Proposition 6.5 (Existence of Exponential Moments of ID Loss Random Variables) *Consider a log-subadditive function* $g(x)$ *that satisfies*

$$g(x+y) \leq Kg(x)g(y), \quad \forall x, y \in \mathbb{R}, \tag{6.43}$$

with $\sup_{\|x\|\leq r} g(x) < \infty$ *for all* $r > 0$*. Then for a loss random variable* X *which is ID with the Lévy–Khinchin triplet given by* $(a, b, W(dx))$*, one has the following moment conditions*

$$\mathbb{E}[g(X)] < \infty \iff \int_{\|x\|>1} g(x)W(dx) < \infty. \tag{6.44}$$

Furthermore, one can state the following bounds on the absolute moments of a loss random variable that is ID as given in Propositions 6.6 and 6.7, see Rosiński & Samorodnitsky (1993), Marcus & Rosiński (2005) and Steutel & Van Harn (2003).

Proposition 6.6 (Bounds on ID Loss Random Variables Absolute Moment) *If* X *is an ID loss random variable with the Lévy triplet* $(\sigma, \gamma, W(dx))$ *and* $l = l(W(dx))$ *is the unique solution to the equation*

$$\int_{\mathbb{R}} \|l^{-1}x\|^2 \wedge \|l^{-1}x\| W(dx) = 1, \tag{6.45}$$

then the following moment bound applies:

$$\frac{1}{4}l(W) \leq \mathbb{E}[|X|] \leq (2.125)l(W). \tag{6.46}$$

In addition, one has the following bound on fractional powers of the absolute first moment of an ID loss random variable.

Proposition 6.7 (Bounds on ID Loss Fractional Powers of Absolute Moment) *If X is an ID loss random variable with the Lévy triplet $(\sigma, \gamma, W(dx))$ and $l := l(W(dx))$ is the unique solution to the equation*

$$l^{-2}\int_{|x|\leq l} x^2 W(dx) + l^{-p}\int_{|x|>l} |x|^p W(dx) = 1, \tag{6.47}$$

for any $p \geq 1$, then the following moment bound applies:

$$\frac{1}{4}l(W) \leq (\mathbb{E}[|X|])^{1/p} \leq K(p)l(W). \tag{6.48}$$

If $p \in (1, 2)$, then $K(p) \leq 2.3$.

One can also state the following large deviations result for ID distributions given in Rosiński & Samorodnitsky (1993) which shows the following result for tail probabilities of ID loss random variables with respect to their Lévy measure.

Proposition 6.8 (Large Deviations for ID Loss Random Variables) *If X is an ID loss random variable with the Lévy triplet $(\sigma, \gamma, W(dx))$ with $W \neq 0$ and taking bounded support on $\mathbb{R}$. Then defining R by the following*

$$R := \inf\{r > 0 : W\{x : |x| > r\} = 0\}, \tag{6.49}$$

one has the following large deviations result

$$\lim_{t\to\infty} \frac{\ln \mathbb{P}\mathrm{r}[|X| > t]}{t \ln t} = -\frac{1}{R}. \tag{6.50}$$

Remark 6.3 *In the case that the Lévy measure W has unbounded support, then this limit is zero.*

The following additional large deviation inequalities were expressed in Houdré (2002) for ID random variables, see Proposition 6.9. Before presenting this result, it is useful to recall the definition of a Lipschitz function given in Definition 6.10.

Definition 6.10 (Lipschitz Function) *A function f is a Lipschitz function if it satisfies*

$$|f(x) - f(y)| \leq C|x - y| \tag{6.51}$$

for all x and y, where C is a constant independent of x and y.

Given this definition, one can state the following large deviations result for ID loss random variables with regard to the properties of their Lévy measure.

Proposition 6.9 (Large Deviation Inequalities) *If X is an ID loss random variable with the Lévy triplet $(0, \gamma, W(dx))$ with the Lévy measure W of bounded support. Then defining R by the following*

$$R := \inf\{r > 0 : W\{x : |x| > r\} = 0\} \tag{6.52}$$

and V according to

$$V^2 = \int_{\mathbb{R}} |x| W(dx), \tag{6.53}$$

one has the following large deviations result for every Lipschitz function $f : \mathbb{R} \mapsto \mathbb{R}$ with $\|f\|_{Lip} \leq 1$ and $t > 0$

$$\mathbb{P}\mathrm{r}[f(X) - \mathbb{E}[f(X)] \geq t] \leq \exp\left(\frac{t}{R} - \left(\frac{t}{R} + \frac{V^2}{R}\right) \ln\left(1 + \frac{Rt}{V^2}\right)\right). \tag{6.54}$$

Furthermore, this implies that for every value $p < \frac{1}{R}$, one has bounded exponential moments for such ID loss random variables given by

$$\mathbb{E}[\exp(p|f(X)| \ln |f(X)|)] < \infty. \tag{6.55}$$

To complete the fundamental understanding of the relationships here between convolutional semi-groups, generators of such groups and the family of the Lévy processes of relevance to OpRisk LDA models, we consider a special case of the Lévy–Ito decomposition of a Lévy process. The case we present is the compound Poisson process where we define the characteristic function for such a process with respect to the Lévy measure. Then we obtain the so-called 'Poisson Jump' measure of the compound Poisson process and use this to define (as per Theorem 6.4) an equivalent representation via the Lévy–Ito decomposition of a Lévy process into components with respect to Poisson processes which we have previously obtained the generators for and characterized earlier.

Definition 6.11 (Characterizing Compound Poisson Lévy Processes) *Let $\{X_t\}_{t \geq 0}$ be a compound Poisson process with jump intensity λ and jump size distribution F on $\mathbb{R}$, represented by*

$$X_t = \sum_{i=1}^{N_t} Y_i(\tau_i), \tag{6.56}$$

where $N_t \sim Poisson\,(\lambda t)$ and i.i.d. $Y_i \sim F_Y(y)$ and the compound Poisson Lévy process has jumps at times $\{\tau_i\}_{i=1:N_t}$ in time increment $[0, t]$. Furthermore, in an OpRisk setting, assume the characteristic function of the losses Y_i is denoted by Φ_Y. Then characteristic function of the compound Poisson process with respect to the Lévy measure is given by

$$\begin{aligned}
\mathbb{E}[\exp(\mathrm{i}\theta X_t)] &= \mathbb{E}[\mathbb{E}[\exp(\mathrm{i}\theta X_t)]|N_t] \\
&= \sum_{n=0}^{\infty} \frac{(\lambda t)^n (\Phi_Y(\theta))^n}{n!} \exp(-\lambda t) \\
&= \exp(\lambda t(\Phi_Y(\theta) - 1)) \\
&= \exp\left(t\lambda \int_{-\infty}^{\infty} (\exp(\mathrm{i}\theta x - 1)\, dF(x)\right)
\end{aligned}$$

$$= \exp\left(t \int_{-\infty}^{\infty} (\exp(\mathrm{i}\theta x - 1)\, W(dx) \right), \tag{6.57}$$

with $\int W(dx) = \lambda$. *Furthermore, one can define the Poisson random measure known as the 'Poisson jump' measure* J_x *of this compound process for any measurable sets* $B = ([t_1, t_2] \times A) \subset [0, \infty) \times \mathbb{R}$ *according to the Lévy measure as* $J_x(B) = W(dx)\,dt = \lambda F(dx)\,dt$. *Then owing to the equivalences detailed in Theorem 6.4 and this definition of the Poisson jump measure, one can utilize the following result (see Cont & Tankov (2004, proposition 3.6)) to show that there is a one-to-one correspondence between compound Poisson processes and Poisson random measures. Given the Poisson random measure* J_X *with intensity measure* $W(dx)\,dt = \lambda F(dx)\,dt$ *that satisfies for all measurable sets* B *the condition that* $\nu(A)[t_2 - t_1] < \infty$ *and given functions* f *satisfying* $\int_B \exp(f(x)) J_X(dt \times dx) < \infty$, *one can show that*

$$\mathbb{E}\left[\int_B f(x) J_X(dt \times dx) \right] = \exp\left(\int_B (\exp(f(x)) - 1)\, W(dx) \right). \tag{6.58}$$

Hence, one can show as a consequence that all compound Poisson processes can be represented in the following equivalent forms

$$X_t = \sum_{i=1}^{N_t} (X_{\tau_i} - X_{\tau_{i-1}}) = \int_{[0,t]\times\mathbb{R}} x J_x(dt \times dx). \tag{6.59}$$

Note that this special representation of the Lévy–Ito decomposition of a Lévy process (see Cont & Tankov (2004, proposition 3.7, p. 79)). To complete the characterization of compound Poisson processes according to its specification via convolutional semi-groups as specified uniquely via a generator, we note that the generator of a Poisson process is given by $\mathcal{G} = \lambda(\mathcal{F} - \mathbf{1})$ *where* $\mathcal{F}$ *denoted the convolution operator with respect to jump distribution* F.

Now given these formal specifications of the compound Poisson process as a Lévy process, in Lemma 6.8 the practically more relevant for OpRisk settings, Poisson compound distribution is obtained through sampling of the Lévy process (annually).

Lemma 6.8 (Compound Poisson Lévy Process to Compound Poisson Distributions) *Given a compound Poisson Lévy process* $\{X_t\}_{t\geq 0}$ *with jump intensity* λ *and jump size distribution* F *on* $\mathbb{R}$, *represented by*

$$X_t = \sum_{i=1}^{N_t} Y_i(\tau_i), \tag{6.60}$$

where $N_t \sim Poisson\,(\lambda t)$ *and i.i.d.* $Y_i \sim F_Y(y)$ *and the compound Poisson Lévy process has jumps at times* $\{\tau_i\}_{i=1:N_t}$ *in time increment* $[0, t]$. *To obtain a representation of the LDA model compound Poisson distribution, one utilizes the independent increments property of a Lévy process and constructs the following annually sampled random variables* $\{Z_{1:T}\}$ *for years* $t \in \{1, 2, \ldots, T\} \in [0, t]$. *Here the annual total loss in year* t *given by* Z_t *is the difference between the Lévy process of the previous year* $(t - 1)$ *and the current year according to* $Z_t = X_t - X_{t-1}$. *This is the same as accumulating all the jumps in a given annual interval according to*

$$Z_t = \sum_{i=1}^{N_t} Y_i(\tau_i) - \sum_{i=1}^{N_{t-1}} Y_i(\tau_i) = \sum_{\tau_i \in [t-1,t]} Y_i(\tau_i). \tag{6.61}$$

Furthermore, the following simple relationship between the probability distribution at a sample time t and the Lévy measure can be obtained,

$$\begin{aligned}\mathbb{P}\mathrm{r}[X_t \in A] &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \exp(-\lambda t)\mathbb{P}\mathrm{r}[X_t \in A | N_t = n] \\ &= \exp(-\lambda t)\mathbb{I}[N_t = 0] + \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} \exp(-\lambda t) F^{(n)*}(A). \end{aligned} \tag{6.62}$$

Then if the jump size distribution (severity distribution) $F(x)$ admits a density $f(x)$ with respect to Lebesgue measure, one obtains

$$\mathbb{P}\mathrm{r}[X_t \in A] = \exp(-\lambda t)\mathbb{I}[N_t = 0] + \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \int_A f^{(n)*}(x)\, dx. \tag{6.63}$$

Finally, to relate this to the Lévy density $\nu(x)$, one can observe that as the time increment goes to zero

$$\lim_{t \downarrow 0} \frac{1}{t} \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} \exp(-\lambda t) f_X^{(n)*}(x) \to \lambda f_X(x) = \nu(x), \ \forall x \neq 0. \tag{6.64}$$

So it is easy to see that the LDA model is trivially recovered from the Lévy process by sampling and differencing the process annually. Two important examples in heavy-tailed risk modeling will be considered based on the NIG-Lévy process and the stable-Lévy process and their compound Poisson Lévy process extensions. First, it is clear that a discussion on the characterization of the distributional properties of samples of a Lévy process is required, as detailed in Lemma 6.9 and Cont & Tankov (2004, section 3.6) and Orey (1968).

Lemma 6.9 (Existence and Relationship between the Lévy Densities and Probability Densities) *If $\{X_t\}_{t \geq 0}$ is a Lévy process with the Lévy triplet $(\sigma, \gamma, W(dx))$ and if the Lévy measure satisfies the condition that there exists a constant $\beta \in (0, 2)$ such that*

$$\liminf_{\varepsilon \downarrow 0} \varepsilon^{-\beta} \int_{-\varepsilon}^{\varepsilon} |x|^2 W(dx) > 0, \tag{6.65}$$

then for any sample time t of the Lévy process, the random variable X_t has a smooth density $f_{X_t}(x)$ such that as $|x| \to \infty$, one has

$$\frac{\partial^n f_{X_t}(x)}{\partial x^n} \to 0, \ \forall n \geq 1. \tag{6.66}$$

Furthermore, the density $f_{X_t}(x)$ is continuous if $\sigma > 0$ or $W(\mathbb{R}) = \infty$.

In OpRisk settings in which we wish to obtain a closed-form representation of the LDA compound distributions sampled from the Lévy process annually, one can clearly observe that this will be possible for the Lévy processes only in general when the sub-class of compound Poisson-ID Lévy processes is considered. That is, compound Poisson Lévy processes in which the jump distribution is ID, because this will correspond to cases in which the closure under convolution property will be of direct relevance for specification of distribution $F^{(n)*}(x)$ and density $f^{(n)*}(x)$. Next we present two special examples, the first is based on the extension of the NIG-Lévy process which has been explored in financial contexts in Barndorff-Nielsen (1997a), Barndorff-Nielsen (1997b), Kalemanova *et al.* (2007) and Rydberg (1997). The second model considered is the stable-Lévy process discussed in Feller (1966).

EXAMPLE 6.5 NIG-Lévy Processes to the Compound Poisson Version

The NIG-Lévy process was studied in Rydberg (1997) where it was presented according to the Lévy triplet $(\sigma, \gamma, W(dx))$ and the Lévy–Khinchine characteristic function given by

$$\mathbb{E}[\exp(\mathrm{i}\theta Y_t)] = \exp(t\psi(\theta)),\ \theta \in \mathbb{R} \tag{6.67}$$

with pure jump process NIG-Lévy characteristic exponent given by

$$\begin{aligned}\psi(\theta) &= \mathrm{i}\gamma\theta + \int_{\mathbb{R}} (\exp(\mathrm{i}\theta y) - 1 - \mathrm{i}\theta y \mathbb{I}[|y| \leq 1]) W(dy) \\ &= \mathrm{i}\gamma\theta + \int_{|y|\geq 1} (1 - \exp(\mathrm{i}\theta y)) f(y; \alpha, \beta, \delta)\, dy \\ &\quad + \int_{|y|<1} (1 - \exp(\mathrm{i}\theta y) - \mathrm{i}\theta y) f(y; \alpha, \beta, \delta)\, dy \end{aligned} \tag{6.68}$$

and values for the density, γ and Lévy measure W are given by

$$\begin{aligned} f(y; \alpha, \beta, \delta) &= \frac{\delta\alpha}{\pi|y|} \exp(\beta y) K_1(\alpha|y|), \\ \gamma &= \frac{2\delta\alpha}{\pi} \int_0^1 \sinh(\beta y) K_1(\alpha y)\, dy, \\ W(dy) &= f(y; \alpha, \beta, \delta)\, dy. \end{aligned} \tag{6.69}$$

To extend this characterization to the compound Poisson NIG Lévy process model, we consider the process given by

$$X_t = \sum_{i=1}^{N_t} Y_i, \tag{6.70}$$

where N_t follows a Poisson process with intensity parameter λ, independent from i.i.d. 'jump sizes' $Y_i \sim F_Y(y) = NIG(y; \alpha, \beta, \delta, \mu)$ which follow an NIG ID distribution. According to Definition 6.11, one, therefore, obtains the characteristic function for the compound Poisson-NIG Lévy process as

$$\begin{aligned} \mathbb{E}[\exp(\mathrm{i}\theta X_t)] &= \exp(\lambda t(\Phi_Y(\theta) - 1)) \\ &= \exp\left(t\lambda \int_{-\infty}^{\infty} (\exp(\mathrm{i}\theta x - 1)\, dF(x) \right), \end{aligned} \tag{6.71}$$

where $\Phi_Y(\theta)$ is the characteristic function of the NIG model and F is the distribution of the NIG jump size model, and because the NIG distribution is closed under convolution, one has that $F^{(n)*}$ is still an NIG distribution for all $n \in \mathbb{J}^+$ such that

$$\begin{aligned} F^{(n)*}(x) = \int_{-\infty}^{x} & \frac{\alpha\widetilde{\delta}_n \mathrm{K}_1(\alpha\sqrt{\widetilde{\delta}_n^2 + (x - \widetilde{\mu}_n)^2})}{\pi\sqrt{\widetilde{\delta}_n^2 + (x-\mu)^2}} \\ & \times \exp(\widetilde{\delta}_n\sqrt{\alpha^2 - \beta^2} + \beta(x - \widetilde{\mu}_n))\, dx \end{aligned} \tag{6.72}$$

with $\widetilde{\delta}_n = n\delta$ and $\widetilde{\mu}_n = n\mu$. This integral is easily evaluated via Fourier transform methods or closed-form series expansions. Now, when this process $\{X_t\}_{t\geq 0}$ is sub-sampled at times $t \in \{1, 2, \ldots, T\}$ which correspond in OpRisk to annual intervals, this will result in the compound Poisson-NIG LDA model.

Simulate a compound Poisson Normal inverse Gaussian Lévy process over [0,*T*]:

1. Generate the number of jumps in interval $[0, T]$ from the homogeneous Poisson process $N_T \sim Poisson(\lambda T)$.
2. Generate the jump times in the interval $[0, T]$ by observing that $\tau_{1:N_T}|N_T \sim Uniform(0, T)$.
3. Generate the jump sizes $Y_i \sim NIG(y; \alpha, \beta, \delta, \mu)$ for all $i \in \{1, 2, \ldots, N_t\}$. This is achieved for each jump Y_i by the following stages of simulation detailed in Rydberg (1997).
   - Generate a standard Normal random variate $V \sim Normal(0, 1)$ and square it to get $P = V^2$.
   - Let $a = \delta$ and $b = \sqrt{\alpha^2 - \beta^2}$, then set $W = \frac{a}{b} + \frac{P}{2b^2} + \frac{\sqrt{4abP + P^2}}{2b^2}$.
   - Generate a uniform random variate $U \sim Uniform[0, 1]$ and perform a rejection stage
   $$Z = \begin{cases} W, & \text{if } U \leq \dfrac{a}{a + Wb}, \\ \dfrac{a^2}{b^2 W}, & \text{otherwise.} \end{cases} \tag{6.73}$$
   - Generate a random Normal variate $Y|Z = z \sim Normal(\mu + \beta z, z)$ which will be unconditionally drawn from the distribution $NIG(y; \alpha, \beta, \delta, \mu)$.
4. Construct the process $\{X_t\}_{t\in[0,T]}$ using the draws of the jump times $\tau_{1:N_t}$ and the jump sizes $Y_{1:N_T}(\tau_{1:N_T})$. ■

EXAMPLE 6.6 α-Stable Lévy Process to the Compound Poisson Version

The semi-group properties of the ID α-stable distribution were first specified in Feller (1966, section IX.5, p. 295, equation 5.14), in which it was shown that if one considers the α-stable convolution semi-group $\{\mathscr{D}(m, F)\}$ in which the distribution of interest is the α-stable distribution scaled by continuous in m constants $\lambda_m > 0$ and translation parameters $\beta_m \in \mathbb{R}$ such that $F(x; m) = F(\lambda_m(x - \beta_m))$, then the generator of the α-stable convolutional semi-group can be obtained. This involves utilizing results which state that continuous convolutional semi-groups can be described in terms of functions ψ^+ and ψ^- defined for the right- and left-half axis according to the relationships as $m \to 0$

$$\frac{1}{m}(1 - F(x; m)) \to \varphi^+(x), \quad \frac{1}{m}F(-x; m) \to \varphi^-(-x), \tag{6.74}$$

at all points of continuity (except a possible atom at the origin). As a result, one can show that for the α-stable process one obtains when the translation parameter $\beta_m = 0$ that the definition of regular variation is obtained for $c^+ \geq 0$

$$\varphi^+(x) = c^+ x^{-\alpha}. \tag{6.75}$$

In addition, Feller (1966) derives from this the generator of the semi-group as

$$\mathscr{G} = \mathscr{G}_\alpha + \gamma \frac{d^2}{dx^2} + \beta \frac{d}{dx} \tag{6.76}$$

with $\mathscr{G}_\alpha = c^+ \mathscr{G}_\alpha^+ + c^- \mathscr{G}_\alpha^-$ and

$$\mathscr{G}_\alpha^+ \phi(x) = \begin{cases} \alpha \int_0^\infty \dfrac{\phi(x-y) - \phi(x)}{y^{\alpha+1}} dy, & \text{if } 0 < \alpha < 1 \\ \alpha \int_0^\infty \dfrac{\phi(x-y) - \phi(x) + T_1(y) \frac{d}{dx}\phi(x)}{y^{\alpha+1}} dy, & \text{if } \alpha = 1 \\ \alpha \int_0^\infty \dfrac{\phi(x-y) - \phi(x) + y \frac{d}{dx}\phi(x)}{y^{\alpha+1}} dy, & \text{if } 1 < \alpha < 2. \end{cases} \tag{6.77}$$

Equivalent relationships hold for the generator $\mathscr{G}_\alpha^-$. Furthermore, it was shown in Sato (1999, chapter 3, section 14) and in Cont & Tankov (2004) that a distribution is α-stable with index of stability $0 < \alpha < 2$ if and only if it is ID with characteristic Lévy triplet given by (σ, γ, W) such that $\sigma = 0$ and

$$W(x) = \frac{A}{x^{\alpha+1}} \mathbb{I}[x > 0] + \frac{B}{|x|^{\alpha+1}} \mathbb{I}[x < 0] \tag{6.78}$$

for constants $A > 0$ and $B > 0$. Note that this shows the relationship between the functions $\varphi^+(x)$ in the convolution semi-group generators and the Lévy measure, as discussed in Proposition 6.4.

To extend this characterization to the compound Poisson α-stable Lévy process model, we consider the process again given by

$$X_t = \sum_{i=1}^{N_t} Y_i, \tag{6.79}$$

where N_t follows a Poisson process with intensity parameter λ, independent from i.i.d. 'jump sizes' $Y_i \sim F_Y(y) = S_\alpha(y; \beta, \gamma, \delta; S(0))$ which follow an α-stable ID distribution. According to Definition 6.11, one, therefore, obtains the characteristic function for the compound Poisson-α-stable Lévy process as

$$\begin{aligned} \mathbb{E}[\exp(\mathrm{i}\theta X_t)] &= \exp(\lambda t(\Phi_Y(\theta) - 1)) \\ &= \exp\left(t\lambda \int_{-\infty}^{\infty} (\exp(\mathrm{i}\theta x - 1)\, dF(x) \right), \end{aligned} \tag{6.80}$$

where $\Phi_Y(\theta)$ is the characteristic function of the α-stable model and F is the distribution of the α-stable jump size model and because the α-stable distribution is closed under convolution, one has that $F^{(n)*}$ is still an α-stable distribution for all $n \in \mathbb{J}^+$.

Now, when this process $\{X_t\}_{t\geq 0}$ is sub-sampled at times $t \in \{1, 2, \ldots, T\}$ which correspond in OpRisk to annual intervals, this will result in the compound Poisson-α-stable LDA model which is simulated as follows:

To simulate a compound Poisson α-stable Lévy process over [0, T]*:*

1. Generate the number of jumps in interval $[0, T]$ from the homogeneous Poisson process $N_T \sim Poisson(\lambda T)$.
2. Generate the jump times in the interval $[0, T]$ by observing that $\tau_{1:N_T}|N_T \sim Uniform(0, 1)$.
3. Generate the jump sizes $Y_i \sim S_\alpha(y; \beta, \gamma, \delta; S(0))$ for all $i \in \{1, 2, \ldots, N_t\}$. This is achieved for each jump Y_i by the following stages of simulation detailed in Chapter 4.
4. Construct the process $\{X_t\}_{t\in[0,T]}$ using the draws of the jump times $\tau_{1:N_t}$ and the jump sizes $Y_{1:N_T}(\tau_{1:N_T})$.

■

Remark 6.4 *It should be noted that in the above two examples when utilizing these models in a practical OpRisk modeling scenario, one should be careful to manage the fact that the support of such distributions may be over the entire real line. A simple truncation to positive support of the severity distributions (jump distributions) can be performed; however, this would defeat the property of 'doubly ID Lévy processes'. This was required to get closure under convolution for such models. In general, this can be addressed through other means and is discussed in the practical section to follow in which either parameter restrictions are imposed to enforce positive support for the severity distribution or a special version of truncation is applied.*

An example of such a process which requires less restrictions, although will not produce the heavy-tailed properties of a sub-exponential models presented under the α-stable and NIG models, is the related family of ID IG jump distributions, discussed in Example 6.3. From a technical perspective, one can also discuss the truncation of the jump distributions in a Lévy process. This relates to the class of TS and truncated Lévy process models, see discussions in detail in Chapter 4 and Cont & Tankov (2004, p. 83).

6.2.2 DOMAIN OF ATTRACTION OF LÉVY PROCESSES: STABLE AND TWEEDIE CONVERGENCE

To conclude these more theoretical discussions, it is appropriate to discuss the domain of convergence of such Lévy processes, as this will motivate the proceeding practical models developed from such results. So we consider the Lévy process $\{X_t\}$ with independent and stationary increments for all t with $X_0 = 0$. Now with such processes in mind, one can consider two types of asymptotic weak convergence results, those based on the large sample asymptotics that result in the α-stable limit theorems discussed in Chapter 4 and those based on ID asymptotics, see discussions in Jørgensen (1991). To clarify the difference between these two asymptotic weak convergence results, we consider examples of each, starting with the more familiar large sample case. Consider the α-stable Lévy process $\{X_t\}_{t>0}$ with all finite-dimensional distributions given by the stable distribution with tail index parameter $\alpha \in (0, 2]$. Then we have seen in Chapter 4 that such a process will satisfy the following stability condition between any

two time increments asymptotically as the increments considered separate infinitely far apart, characterized by the statement

$$r^{-\frac{1}{\alpha}} X_{rt} \stackrel{d}{=} X_t, \quad \forall t \geq 0. \tag{6.81}$$

As discussed in Jørgensen (1991), this α-stable stability condition can now be utilized to characterize the behaviour and domains of attraction of other Lévy processes under two regimes:

- large sample asymptotics with $r \to \infty$ that result in domains of attraction of the Lévy processes that will converge to α-stable limits; and
- small sample asymptotics with $r \to 0$ that have been termed the infinitely divisible asymptotics.

For example, the large sample asymptotic of any Lévy process $\{Y_t\}$ that will have a limiting form given by an α-stable Lévy process $\{X_t\}$ are characterized by the stable domain of attraction. A Lévy process Y_t is in the domain of attraction of a stable Lévy process X_t if it satisfies the large sample asymptotic given by the scaling term

$$\lim_{r \to \infty} r^{\frac{1}{\alpha}} Y_{rt} \xrightarrow{d} X_t, \quad \forall t > 0. \tag{6.82}$$

This can be stated more formally with the additional incorporation of drift components, which in OpRisk could allow one to incorporate factors such as key risk indicators that may affect the loss process and can be incorporated into the Lévy process model as a regression drift. The formal statement of this result is presented in Proposition 6.10, see discussions in Jørgensen (1991).

Proposition 6.10 (Weak Convergence of Lévy Processes to Stable Lévy Processes) *Given the family of stable-Lévy processes given by $S_t(\mu, \alpha) = S_t + \mu t$ for drift $\mu \in \mathbb{R}$ and stability index $\alpha \in (0, 2]$. If there exists constant scaling such that a general Lévy process, denoted by $X_t(\mu) = X_t + \mu t$, exists and satisfies one of the two following weak convergence results, either the large sample asymptotic*

$$\lim_{r \to \infty} r^{-\frac{1}{\alpha}} X_{rt} \left(\mu r^{\frac{1}{\alpha} - 1} \right) \xrightarrow{d} S_t(\mu, \alpha), \quad \forall t \geq 0 \tag{6.83}$$

or the small sample asymptotic

$$\lim_{r \downarrow 0} r^{-\frac{1}{\alpha}} X_{rt} \left(\mu r^{\frac{1}{\alpha} - 1} \right) \xrightarrow{d} S_t(\mu, \alpha), \quad \forall t \geq 0, \tag{6.84}$$

then the Lévy process $X_t(\mu)$ is said to be in the domain of attraction of the stable Lévy process $S_t(\mu, \alpha)$ for the given characteristic tail exponent α.

In the case of the limit given in Equation 6.82, we see that there exists non-empty elements in the fixed point set, family of distributions that result from the weak convergence in the small sample asymptotic case as $r \downarrow 0$. In Jørgensen (1991), it is demonstrated that one may also achieve other fixed-point solutions outside of the α-stable Lévy process result discussed earlier, these correspond to considering modifications of natural exponential family (NEF) processes. The NEF process (see discussions and characterizations in Küchler & Sorensen (1997) is given in Definition 6.12 and relates to the tilted processes discussed in Chapter 4. This family of

Lévy processes is known as the 'natural exponential family' (NEF) of Lévy processes generated by Lévy process $\{X_t\}_{t>0}$.

Definition 6.12 (Natural Exponential Family Lévy Processes) *An NEF-Lévy process on a probability space $(\mathcal{X}, \mathcal{F}_t, \mathbb{P}(\theta))$ is characterized by a generating Lévy process $\{X_t\}_{t>0}$ on a probability space $(\mathcal{X}, \mathcal{F}, \mathbb{P})$ and a cumulant-generating function*

$$\kappa(\theta) = \ln\mathbb{E}[\exp(\theta X_1)] \tag{6.85}$$

with domain $\theta \in \Theta = \{\theta \in \mathbb{R} : \kappa(\theta) < \infty\}$ such that $\Theta \neq \{0\}$. Then for all $t \geq 0$, one defines the probability measures $\{\mathbb{P}_t(\theta) : \theta \in \Theta\}$ on the space $(\mathcal{X}, \mathcal{F}_t)$ by the Radon–Nikodym derivative for any θ w.r.t. the base-generating Lévy processes measure and the cumulant function by

$$\frac{d\mathbb{P}_t(\theta)}{d\mathbb{P}_t} = \exp(\theta X_t - t\kappa(\theta)), \tag{6.86}$$

with $\mathbb{P}_t$ the finite-dimensional marginal distribution of X_t with filtration $\mathcal{F}_t$. The corresponding restriction of the family of measures $\mathbb{P}(\theta)$ at any time t to the filtration $\mathcal{F}_t$ produces a Lévy process on the space $(\mathcal{X}, \mathcal{F}, \mathbb{P}(\theta))$.

This generation mechanism is known as the natural exponential tilting of a generating Lévy process, see discussions in the stable Lévy process setting in Chapter 4. Here the NEF-Lévy measure $\mathbb{P}(\theta)$ is obtained by the exponential tilting of the base Lévy processes measure $\mathbb{P}$ (of course as discussed in Chapter 4); this is equivalent to the tempering of the Lévy spectral measure.

As in the case of the exponential family of models in generalized linear model (GLM) regression analysis, one can also characterize the NEF-Lévy processes by the mean and variance functions as detailed in Definition 6.13.

Definition 6.13 (NEF-Lévy Process Characterization via Mean and Variance Function) *Given a NEF-Lévy process $\{X_t\}_{t>0}$ characterized with respect to the cumulant function*

$$\kappa(\theta) = \ln\mathbb{E}[\exp(\theta X_1)]. \tag{6.87}$$

Then for all t, one has the marginal measure $\mathbb{P}^t(\theta)$ which corresponds to an NEF measure on $\mathbb{R}$ which can be characterized by the mean and variance functions given by

$$\begin{aligned}\mathbb{E}[X_t] &= t\frac{d}{d\theta}\kappa(\theta) = t\mu,\\ \mathbb{V}\mathrm{ar}[X_t] &= t\frac{d^2}{d\theta^2}\kappa(\theta) = tV(\mu).\end{aligned} \tag{6.88}$$

Remark 6.5 *One can then characterize many different choices of variance functions, one popular choice is the power (monomial) variance functions $V(\mu) = \mu^p$ that were utilized in Lee & Whitmore (1993) and in the Poisson-Tweedie sub-class of models in Peters* et al. *(2009b). The power variance functions when utilized with a Lévy process generator to create a sub-family of NEF-Lévy processes are often termed 'Hougaard-Lévy processes' for $p \in \mathbb{R} \notin (0, 1)$. The domain of the Hougaard-Lévy processes behaves differently for special points, such as when $p = 0$, one*

obtains that the process $\{X_t\}$ *takes values in the whole real line, whereas when values of* $p \in \mathbb{R} \notin \{\{0\}, (1, 2)\}$, *the Hougaard-Lévy processes only takes positive support* $X_t \in \mathbb{R}^+$.

Having defined the NEF-Lévy process family, one can define formally the special sub-family of Hougaard-Lévy processes as given in Definition 6.14, see Lee & Whitmore (1993) and the special cases in Peters *et al.* (2009b).

Definition 6.14 (Hougaard-Lévy Processes) *A Lévy process* $\{X_t\}_{t>0}$ *is called a Hougaard-Lévy process, denoted by* $\{H_t(\alpha, \eta, \kappa)\}_{t>0}$ *if the following Lévy process characteristics are satisfied:*

1. $X_0 = 0$*; and*
2. $\{X_t\}$ *has stationary-independent increments; and in addition one has*
3. $\{X_t\}$ *has the Laplace transform given by*

$$\mathbb{E}[\exp(-\theta X_t)] = \exp(-t\zeta[(\kappa + \theta)^\alpha - \kappa^\alpha]) \tag{6.89}$$

for $\alpha \in (0, 1)$, $\zeta > 0$ *and* $\kappa \geq 0$.

The Lévy characteristic function representation is then given at any time $t > 0$ *by the representation*

$$\begin{aligned} \Phi_{X_t}(\theta) &= \mathbb{E}[-\theta X_t] \\ &= \exp(-t\zeta[(\kappa + \theta)^\alpha - \kappa^\alpha]) \\ &= \exp\left(t \int_0^\infty (1 - \exp(-\theta x))\, W(dx)\right) \end{aligned} \tag{6.90}$$

with the Lévy measure for the Hougaard-Lévy process given by

$$W(dx) = \frac{\alpha\zeta}{\Gamma(1-\alpha)} \int_x^\infty u^{-1-\alpha} \exp(-\kappa u)\, du. \tag{6.91}$$

It is well known that the Hougaard-Lévy process can generate three well-known NEF-Lévy processes given by

1. $\kappa = 0$, then the process $\{H_t(\alpha, \eta, \kappa = 0)\}_{t>0}$ is a positive support α-stable Lévy process, see Weron (1984);
2. $\alpha = \frac{1}{2}$, then the process $\left\{H_t(\alpha = \frac{1}{2}, \eta, \kappa)\right\}_{t>0}$ is an IG Lévy process, see Chhikara & Folks (1989); and
3. $\alpha \to 0$ and $\kappa > 0$, then the process $\{H_t(\alpha \to 0, \eta, \kappa > 0)\}_{t>0}$ one obtains is the gamma Lévy process.

Remark 6.6 *As discussed in Jørgensen (1991), the Hougaard-Lévy process* $\{H_t(\alpha, \zeta, \kappa)\}$ *can also be re-expressed in terms of the NEF-Lévy process as those NEF-Lévy processes that have a variance function given by the power transform* $V(\mu) = \mu^p$ *denoted under this characterization by* $\{H_t(p, \mu)\}$.

Given this NEF-Lévy process characterization of the power-transformed variance function NEF-Lévy processes to form the Hougaard-Lévy family, one can show that $\{H_t(p, \mu)\}$ form

the fixed point of the normalized Hougaard-Lévy processes in which for all constants $r > 0$ one has weak equality (in distribution) for any time $t > 0$ given for any $p \in \mathbb{R} \notin (0, 1)$ by

$$r^{-\frac{1-p}{2-p}} H_{rt}\left(p, r^{\frac{1-p}{2-p}-1}\mu\right) \stackrel{d}{=} H_t(p, \mu). \tag{6.92}$$

One can then state the following results regarding the large sample and small sample asymptotics of the NEF-Lévy family of processes under scaling relative to the Hougaard fixed point as detailed in Proposition 6.11, see discussions in Tweedie (1947), Wedderburn (1974), Jørgensen (1991) and Jørgensen *et al.* (1994). The results on the Tweedie convergence as characterized by the convergence of variance functions can be found in Letac (1989) and Mora (1990) for ED models.

Proposition 6.11 (Tweedie Convergence and Hougaard-Lévy Domain of Attraction) *If $\{X_t(\mu)\}_{t>0}$ is a Lévy process with drift μ then it belongs to the domain of attraction of a power variance (Hougaard) Lévy process with variance function $V(\mu) = \mu^p$ for its finite-dimensional distributions if the constant scaling given below satisfies weak convergence in either large sample asymptotics or small sample asymptotics accordig to*

$$\begin{aligned} &\lim_{r\to\infty} r^{-\frac{1-p}{2-p}} X_{rt}\left(r^{\frac{1-p}{2-p}-1}\right) \stackrel{d}{\longrightarrow} H_t(p, \mu), \forall t \geq 0; \text{ or} \\ &\lim_{r\downarrow 0} r^{-\frac{1-p}{2-p}} X_{rt}\left(r^{\frac{1-p}{2-p}-1}\right) \stackrel{d}{\longrightarrow} H_t(p, \mu), \forall t \geq 0. \end{aligned} \tag{6.93}$$

The Tweedie convergence theorem can then be utilized to state that a significant component of the domain of attraction of the Lévy processes that converge to the Hougaard-Lévy fixed point may be characterized by the power asymptotic behaviour of the variance function $V(\mu)$ for the process $\{X_t(\mu)\}_{t>0}$ where

$$V(\mu) \sim \mu^p, \quad as \quad \mu \downarrow 0, \quad or \quad \mu \to \infty. \tag{6.94}$$

In OpRisk settings, there are two particularly relevant examples of Tweedie convergence to be considered when characterizing closed-form LDA models, those corresponding to the Poisson process and the compound Poisson process. We summarize later two examples of relevance to OpRisk discussed in detail in Jørgensen (1991), see Examples 6.7 and 6.8

EXAMPLE 6.7 Large and Small Sample Asymptotics and Tweedie Convergence for Poisson Process

For the Poisson process, one has $p = 1$ and $\alpha = 1 + \frac{1}{(1-p)} \to \infty$, then the Poisson process with $H_t(p = 1, \mu)$ for rate $\mu > 0$ satisfies the fixed-point equation

$$H_{rt}(p = 1, r^{-1}\mu) \stackrel{d}{=} H_t(p = 1, \mu), \quad \forall t \geq 0. \tag{6.95}$$

One can now utilize this property of scaling of Poisson process rates (known as the Poisson Mapping theorem in Kingman (1992)) to characterize the large sample and small sample (infinitely divisible) asymptotics, see Jørgensen (1991).

1. *Large Sample Asymptotic.* In the case of the large sample asymptotic for a Lévy process $\{X_t(\mu)\}_{t>0}$, one has $V(\mu) \sim \mu$ as $\mu \downarrow 0$ which will be satisfied if the distribution of $X_1(\mu)$ takes a positive probability mass at the origin, and in addition, if $(0, 1)$ is the largest interval starting from the origin that has zero probability,

then one has the following Lévy process convergence to the Poisson process

$$\lim_{r\to\infty} X_{rt}(r^{-1}\mu) \xrightarrow{d} H_t(p=1,\mu), \quad \forall t \geq 0 \text{ and } \mu > 0. \tag{6.96}$$

2. *Small Sample (ID) Asymptotic.* In the case of the small sample asymptotic for a Lévy process $\{X_t(\mu)\}_{t>0}$, one has $V(\mu) \sim \mu$ as $\mu \to \infty$ which will be satisfied when one has the following Lévy process convergence to the Poisson process

$$\lim_{r\downarrow 0} X_{rt}(r^{-1}\mu) \xrightarrow{d} H_t(p=1,\mu), \quad \forall t \geq 0 \text{ and } \mu > 0. \tag{6.97}$$

■

EXAMPLE 6.8 Large and Small Sample Asymptotics and Tweedie Convergence for Compound Poisson Process

For the compound Poisson process with $p \in (1,2)$ or $\alpha = 1 + \frac{1}{(1-p)} < 0$, one has the Hougaard-Lévy processes $H_t(p,\mu)$ corresponding to the class of compound Poisson-gamma processes, that is, LDA models comprised compound Poisson-Gamma processes, which is a Poisson process with gamma jumps. Then decompose the drift (mean function) as follows, according to $\mu = \mu_1\mu_2$, and α to get

$$\begin{aligned} \mu_1 &= \frac{\alpha-1}{\alpha}\mu^{\frac{\alpha}{\alpha-1}}, \\ \mu_2 &= \frac{\alpha}{\alpha-1}\mu^{\frac{1}{1-\alpha}}. \end{aligned} \tag{6.98}$$

Now define the Hougaard Lévy process by the compound sum given by

$$H_t(p,\mu) = \sum_{i=1}^{N_t} X_i, \tag{6.99}$$

where $N_t \sim H_t(p=1,\mu_1)$ and the i.i.d. severities given by $X_i \sim Gamma(-\alpha,\beta)$ such that $-\alpha > 0$ and $\mathbb{E}[X_i] = \mu_2$. Then the compound Poisson-Gamma process satisfies the following fixed point

$$H_{rt}(p=1,r^{-1}\mu) \stackrel{d}{=} H_t(p=1,\mu), \quad \forall t \geq 0. \tag{6.100}$$

To characterize the large and small sample asymptotics, one could consider the case of restricting $p = \frac{3}{2}$ ($\alpha = -1$) as in Jørgensen (1991) and Jørgensen (1991) which would produce the class of LDA models obtained from a Poisson-exponential Lévy Process structure. The more general case of the Poisson-Gamma case is given below when we discuss the Tweedie family of LDA models. In the exponential jump case, one has the following two asymptotics:

1. *Large Sample Asymptotics.* In the case of the large sample asymptotic for a Lévy process $\{X_t(\mu)\}_{t>0}$, one has $V(\mu) \sim \mu^{3/2}$ as $\mu \downarrow 0$ which will be satisfied if the

distribution of $X_1(\mu)$ takes a positive probability mass on $\mathbb{R}^+$ with a positive probability mass at 0, then one has the following Lévy process convergence to the compound Poisson-exponential process

$$\lim_{r\to\infty} X_{rt}(r^{-2}\mu) \xrightarrow{d} H_t(p = 3/2, \mu), \quad \forall t \geq 0 \text{ and } \mu \downarrow 0. \tag{6.101}$$

2. *Small Sample (ID) Asymptotics.* In the case of the small sample asymptotic for a Lévy process $\{X_t(\mu)\}_{t>0}$, one has $V(\mu) \sim \mu^{3/2}$ as $\mu \to \infty$ which will be satisfied when one has the following Lévy process convergence to the Poisson process

$$\lim_{r\downarrow 0} X_{rt}(r^{-2}\mu) \xrightarrow{d} H_t\,(p = 3/2, \mu), \quad \forall t \geq 0 \text{ and } \mu \to \infty. \tag{6.102}$$

■

It is typical in practice for OpRisk modeling settings to consider the discrete time versions of these ideas; this leads one to the closed-form LDA models that are studied in the next section, where we consider the Poisson-Tweedie family of LDA models, the Poisson-IG family of LDA models, the Poisson-NIG LDA models and the Poisson-stable LDA models.

6.3 Practical Closed-Form Characterization of Families of LDA Models for Light-Tailed Severities

In this section, we consider the development of LDA models which admit closed-form representations for the annual loss distribution and density functions for general classes of frequency and severity models that will be of direct relevance to applications in practical settings. In particular, we will discuss the models for the light-tailed severity loss processes given by the Gamma, IG and NIG models that were discussed previously in Cruz *et al.* (2015). Here we focus particularly on their closed-form annual loss representations. We begin by defining the ED models and the Poisson-Tweedie models for OpRisk settings which are the discrete time analogues of the special components of the NEF-Lévy processes discussed earlier. The ED models can be considered as special LDA model constructions based on the Esscher transform or exponential tilting that was discussed in the context of pricing in (Cruz *et al.* 2015, chapters 17 and 18). In the statistics literature, a family constructed from an Esscher transform or tilting is typically given the name of the natural exponential families (NEF), see discussion in Jørgensen *et al.* (2009).

6.3.1 GENERAL PROPERTIES OF EXPONENTIAL DISPERSION AND POISSON-TWEEDIE MODELS FOR LDA STRUCTURES

The tilting of a base distribution (tempering of the spectral measure) has been used to create new families of distributions in many contexts throughout the risk and actuarial literature. It has been known as the Esscher transform for insurance pricing and Cramer's transform in large deviations literature, see Ibragimov *et al.* (1971) and Vinogradov (1994). In the statistics literature, the exponential tilting has been shown to create the TS models, and in this section, we will show that it can be utilized effectively to develop families of closed-form LDA models

for OpRisk settings with many practical and interesting features to be considered. Aspects of the tilted families of models have also been developed to describe generalized convergence theorems for sequences of i.i.d. random variables and the characterization of the corresponding domains of attraction for such sequences which are typically known as Tweedie convergence theorems, discussed previously.

The ED models are an important class of models in statistics that have been utilized to generalize the exponential family in many regression contexts, see detailed discussions in Jørgensen (1997). There are several sub-families of such models that have been studied, but typically one distinguishes between two main types of member, the additive ED models and the reproductive ED members. In addition, within the class of ED models, there is a particularly interesting sub-family of models known as the Tweedie models or Poisson-Tweedie models which in the OpRisk context may be utilized to characterize the limiting distributions for closed-form expressions of random sums of independent loss random variables. We begin by presenting the definition of the ED family and then characterize the sub-family of Tweedie models, see detailed discussions in Jørgensen (1987).

We will first present a very general representation for the NEF models, explaining the notion of steepness and infinite divisibility for such models. Then we will provide particular examples for such families of models and how they apply to OpRisk settings. To understand the NEF (tilted distributional families) or ED models, we will first provide a brief introduction to the background, based on general presentation developed in Jørgensen *et al.* (2009). Consider the measure $\nu(dz)$ that takes support on the real line $\mathbb{R}$ and define the coefficient functions with respect to the tilting parameter θ by

$$b_i(\theta) = \int \exp(\theta z) z^i \nu(dz), \quad i = 0, 1, 2. \tag{6.103}$$

Now define the support of the parameter θ by the set referenced with respect to measure ν given by $\Theta_\nu = \{\theta \in \mathbb{R} : b_0(\theta) < \infty\}$, and as discussed in Jørgensen *et al.* (2009), the measure $\nu(dz)$ will be finite only if the parameter θ can take values at the origin $\theta = 0 \in \Theta_\nu$, that is, when one has

$$b_0(\theta) = \int \nu(dz) < \infty. \tag{6.104}$$

Given these definitions, one can then construct the following families of tilted distributions (NEF), with respect to the parameter $\theta \in \Theta_\nu$, according to

$$F_Z(dz; \theta) = \frac{1}{b_0(\theta)} \exp(\theta z) \nu(dz), \tag{6.105}$$

where for each $\theta \in \Theta_\nu$, the distribution satisfies the following:

- the distribution $F_Z(dz; \theta)$ is not degenerate (point mass) for any $\theta \in \Theta_\nu$; and
- the set of valid tilting parameters for base measure ν given by Θ_ν is not degenerate, that is, it must have a cardinality greater than 1.

One can then observe the following important property: given a measure $\nu(dz; \theta)$ with tilting parameter $\theta \in \Theta_\nu$ defined according to $\nu(dz; \theta) = \exp(\theta z)\nu(dz)$, one can generate any other member of the same NEF as those generated by $\nu(dz)$ provided that $\{0\} \in \Theta_\nu$ which implies

that one can consider the following relationship between any members of the NEF generated by either of these measures

$$F_Z(dz;\theta) = \frac{b_0(0)}{b_0(\theta)} \exp(\theta z) F_Z(dz;0). \tag{6.106}$$

This is clearly the Esscher transform discussed in Cruz et al (2015, chapter 18). Two more important components are required to complete the specification, the mean function and variance function of the distribution $F_Z(dz;\theta)$ denoted generically in the literature by the following:

- Mean function of tilted distribution in terms of $b_i(\theta)$ functions is given by

$$\mu = \tau(\theta) := \frac{b_1(\theta)}{b_0(\theta)}. \tag{6.107}$$

  Note that the tilting parameter θ is in a one-to-one correspondence with the mean function mapping $\mu = \tau(\theta)$ which is invariant to exponential tilting of the base-generating measure, so it can be utilized to reparameterize the NEF representation, as shown below.

- Variance function of tilted distribution in terms of $b_i(\theta)$ functions is given by

$$V(\mu) := \tau'(\tau^{-1}(\mu)). \tag{6.108}$$

  This variance function has been studied extensively for NEFs and it has been found to characterize uniquely the NEF that it was constructed from and, therefore, represents all information in the given sub-family of NEFs that will produce a given variance function form. Note that, in this section, we will pay particular attention to the sub-family of power variance functions. This characterization by the variance function of the sub-family of NEFs can be considered as the analogue characterization of a family of distributions by the characteristic function (see Chapter 4 on stable laws for examples).

Further details on the mean and variance functions are available in Jørgensen *et al.* (2009). To complete the discussion on NEFs that has been developed above, it is important to understand the characterization of such models by the notion of steepness and regularity as given in Definition 6.15.

Definition 6.15 (Steepness and Regularity of NEF Models) *Consider the convex support of the distributions $F_Z(dz;\theta)$ for all $\theta \in \Theta_\nu$ with regard to some base-generating measure $\nu(dz)$ denoted by $\mathcal{S}$ which will correspond to the smallest interval of the real line that has a probability of 1. If one then considers the reparameterization of the family of NEFs by the mean function $\mu = \tau(\theta)$ on support $\Omega_\nu := \{\mu : \mu = \tau(\theta), \ \ for \ \ \theta \in interior\ \Theta_\nu\}$ to produce the family $\{F_Z(dz;\mu)\}_{\mu\in\Omega_\nu}$, then an NEF can be characterized by the two following properties (see Jørgensen* et al. *2009):*

- *Regularity of a given NEF sub-family if one has that $\Omega \subseteq int\mathcal{S}$, that is, the set Θ_ν is open and it is called locally regular at boundary points of Θ_ν, that is, at either $\theta_- := \inf \Theta_\nu$ or $\theta_+ := \sup \Theta_\nu$ if either $\theta_- \notin \Theta_\nu$ or $\theta_+ \notin \Theta_\nu$.*
- *Steepness of a given NEF sub-family implies that the family has Ω_ν given by the interior of the convex support of the distribution $F_Z(dz;\theta)$ for a base measure ν, that is, $\Omega_\nu = int\mathcal{S}$, where $int\ \mathcal{S}$ is the interior of $\mathcal{S}$.*

It is now possible to take these general concepts for the definition of a NEF family and to modify them to develop a representation that will produce the ED family of models, see Definition 6.16.

Definition 6.16 (Exponential Dispersion Models) *A random variable $Z \sim ED(\theta, \lambda)$ takes an ED model form with canonical parameter θ and index parameter λ if it has a distribution given by the form*

$$\mathbb{P}r[Z \in A] = \int_A \exp(\theta z - \lambda\kappa(\theta))\nu_\lambda(dz) \tag{6.109}$$

with cumulant function given by

$$\kappa(\theta) = \frac{1}{\lambda} \ln \int \exp(\theta z)\nu_\lambda(dz). \tag{6.110}$$

From the cumulant function choice, one can find the relationship between the canonical parameter θ and the mean and variance functions μ and $V(\mu)$ given in terms of the mapping τ according to

$$\begin{aligned} \mu = \tau(\theta) &= \frac{d}{d\theta}\kappa(\theta), \\ V(\mu) &= \frac{d}{d\theta}\tau(\tau^{-1}(\mu)). \end{aligned} \tag{6.111}$$

The density function is given by

$$f_Z(z; \lambda, \theta) = c(\lambda, z) \exp(\lambda[z\theta - \kappa(\theta)]), \quad z \in \mathbb{R}, \tag{6.112}$$

for given normalization function $c(\cdot)$ and parameters: canonical parameter $\theta \in \Omega$ and index parameter $\lambda \in \Lambda$. The name exponential dispersion (ED) comes about by considering the dispersion parameter $\sigma^2 = \frac{1}{\lambda}$. Note that if X has a density in a sub-family of the ED family in which each sub-member is ID, then $\Lambda = \mathbb{R}^+$.

As was studied in the advanced chapter on severity models for α-stable and tempered α-stable models and also above in the section on the modes of weak convergence termed the infinitely divisible (ID) small sample limit and large sample Tweedie convergence, one can also characterize the Tweedie family within the ED class by the concept of the domain of attraction. We will first introduce the Tweedie models as a special case of the Lévy processes discussed earlier in the context of a compound Process as would be utilized for LDA models in OpRisk.

This family of models is particularly relevant to OpRisk in that they characterize the class of additive exponential models which have the property that the partial sums $Z_n = X_1 + \cdots + X_n$ will have an ED family distribution given by Proposition 6.12.

Proposition 6.12 (Closure Under Convolution of Additive Exponential Dispersion Models) *Given $N = n$ independent loss random variables $X_i \sim ED(\theta, \lambda_i)$, then the partial sum for the annual losses given n losses occurred is given by*

$$Z_n = \sum_{i=1}^{n} X_i \sim ED\left(\theta, \sum_{i=1}^{n} \lambda_i\right). \tag{6.113}$$

One may show that the mean and variance of each additive ED loss random variables is given with respect to the mean and variance functions $\mathbb{E}[X_i] = \lambda_i \mu$ *and* $\mathbb{V}\text{ar}(X_i) = \lambda_i V(\mu)$*; hence, the linear combination will have mean and variance given by*

$$\begin{aligned} \mathbb{E}[Z_n] &= \sum_{i=1}^{n} \lambda_i \mu \\ \mathbb{V}\text{ar}[Z_n] &= \sum_{i=1}^{n} \lambda_i V(\mu). \end{aligned} \tag{6.114}$$

In the case of the reproductive ED family members, one has the following weighted closure under convolution properties, where the reproductive ED family members are defined by taking an ED family members $X \sim ED(\theta, \lambda)$ and scaling it to get a new random variable $Y = \frac{X}{\lambda} \sim ED(\mu, \sigma^2)$ with $\sigma^2 = \frac{1}{\lambda}$. Then the closure of the weighted average of these transformed random variables is given below.

Proposition 6.13 (Closure Under Convolution of Reproductive Exponential Dispersion Models) *Given* $N = n$ *independent loss random variables* $X_i \sim ED(\theta, \lambda_i)$ *which are scaled to give reproductive ED members* $Y_i = \frac{X_i}{\lambda_i} \sim ED\left(\mu, \frac{\sigma^2}{w_i}\right)$ *such that* $\sum_{i=1}^{n} w_i = 1$*, then the partial weighted sum for the annual losses given* n *losses occurred is given by*

$$Z_n = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i Y_i \sim ED\left(\mu, \frac{\sigma^2}{\sum_{i=1}^{n} w_i}\right). \tag{6.115}$$

The sub-class of ED models of interest in this section are those that are both additive and reproductive members of the ED model family, known as the Tweedie distributions. In addition, the Tweedie sub-family of ED models has the property of scale invariance as shown in Proposition 6.14.

Proposition 6.14 (Scale Invariance of ED Family Loss Models) *Given a loss random variable* X *with ED loss model which is scaled to get a reproductive form* $Y \sim ED(\mu, \sigma^2)$*, then the scaled version is also in the ED family given by*

$$cY \sim ED(c\mu, c^{2-p}\sigma^2), \tag{6.116}$$

for index parameter $p \in \mathbb{R}$*,.*

Now using the fact that the Tweedie sub-family of ED models encompasses both additive and reproductive ED features, one can show that the variance function (because of the above-mentioned scale invariance property) must have a functional form which is a power law, see discussion in Jorgensen (1987). Therefore, the Tweedie distributions form a sub-family of probability distributions in that ED class that have the special property for any annual loss random variable Z that has a Tweedie distribution will satisfy the following property for the mean and variance

$$\mathbb{V}\text{ar}[Z] = a(\mathbb{E}[Z])^p, \forall p \in \mathbb{R} \notin (0, 1), \tag{6.117}$$

with a and p given by positive constants. This family of distributions was named in honour of the work performed by Maurice Charles Kenneth Tweedie in characterizing such a distributional family, see Tweedie (1984) and developments in estimation and Bayesian models in Jørgensen & Paes De Souza (1994) and Peters *et al.* (2009b). The parameter p plays a very important role in specification of the Poisson-Tweedie models and their properties as will be discussed. In addition, one may observe that the Tweedie models are related to the discrete time versions of the NEF-Lévy processes with power variance functions discussed previously in this chapter. It is also important to note the following property of the ED family power variance sub-families (Tweedie models) given in Theorem 6.6, see discussion in Jorgensen (1987, theorem 2).

Theorem 6.6 *The power variance functions $V(\mu) \propto \mu^p$ will not produce ED models for powers in the interval $p \in (0, 1)$.*

Keeping this property in mind, one can write the cumulant function of the Tweedie family of ED models as given in Proposition 6.15.

Proposition 6.15 (Tweedie Loss Model Cumulant Function) *Given a loss random variable with a Tweedie model characterized by the cumulant function*

$$\kappa(\theta) = \frac{1}{\lambda} \ln \int \exp(\theta z) \nu_\lambda(dz), \tag{6.118}$$

such that

$$\begin{aligned} \mu = \tau(\theta) &= \frac{d}{d\theta}\kappa(\theta), \\ V(\mu) &= \frac{d}{d\theta}\tau(\tau^{-1}(\mu)) \propto \mu^p. \end{aligned} \tag{6.119}$$

Then the base cumulant function of the Tweedie model can be shown to take one of the following forms, depending on the power index parameter p as follows:

$$\kappa(\theta, p) = \begin{cases} \exp(\theta), & p = 1, \\ -\ln(-\theta), & p = 2, \\ \dfrac{\alpha - 1}{\alpha}\left(\dfrac{\theta}{\alpha - 1}\right)^{\alpha}, & p \neq 1, 2 \quad \text{and} \quad p \geq 1, \end{cases} \tag{6.120}$$

where the Tweedie exponent is given by $\alpha = \frac{p-2}{p-1}$. This base cumulant function can then be utilized to define the cumulant function for the two types of Tweedie family members:

1. Additive Tweedie Loss Models. $X \sim Tweedie(\theta, \lambda, p; A)$ *with characterization via the cumulant function, where the notation A denotes the additive sub-family of Tweedie models:*

$$\ln \mathbb{E}[\exp(tX)] = \kappa(t;\ \lambda, \theta, p;\ A) = \begin{cases} \lambda \exp(\theta)\left[\exp(t) - 1\right], & p = 1, \\ -\lambda \ln\left(1 + \frac{t}{\theta}\right), & p = 2, \\ \lambda\kappa(\theta, p)\left[\left(1 + \frac{t}{\theta}\right)^{\alpha} - 1\right], & p \neq 1, 2. \end{cases} \tag{6.121}$$

2. Reproductive Tweedie Loss Models. $X \sim Tweedie(\theta, \lambda, p; R)$ *with characterization via the cumulant function, where the notation* R *denotes the reproductive sub-family of Tweedie models:*

$$\ln \mathbb{E}[\exp(tX)] = \kappa(t;\ \lambda, \theta, p;\ R) = \begin{cases} \lambda \exp(\theta)\left[\exp\left(\frac{t}{\lambda}\right) - 1\right], & p = 1, \\ -\lambda \ln\left(1 + \frac{t}{\theta\lambda}\right), & p = 2, \\ \lambda\kappa(\theta, p)\left[\left(1 + \frac{t}{\theta\lambda}\right)^{\alpha} - 1\right], & p \neq 1, 2 \\ & \text{and } p \geq 1. \end{cases} \tag{6.122}$$

Within the family of power variance functions (Tweedie model sub-family), one can obtain many important parametric model sub-families characterized by the power parameter p as follows (Jorgensen 1987, table 1):

1. *Extreme Stable-Generated Sub-Family.* If $p < 0$ with $\alpha \in (1, 2)$, then the resulting ED model will be on support $\mathbb{R}$ with $\Omega = \mathbb{R}^+$ and $\Lambda = [0, \infty)$;
2. *Gaussian Sub-Family.* If $p = 0$ with $\alpha = 2$, then the resulting ED model will be a Gaussian model on support $\mathbb{R}$ with $\Omega = \mathbb{R}$ and $\Lambda = \mathbb{R}$;
3. If $p \in (0, 1)$, then there is no ED model;
4. *Poisson Sub-Family.* If $p = 1$ with $\alpha = -\infty$, then the resulting ED model will be a Poisson distribution model on support $\mathbb{N}$ with $\Omega = \mathbb{R}^+$ and $\Lambda = \mathbb{R}$;
5. *Compound Poisson Sub-Family.* If $p \in (1, 2)$ with $\alpha < 0$, then the resulting ED model will be a compound Poisson model on support $\mathbb{R}^+$ with an atom at the origin and with $\Omega = \mathbb{R}^+$ and $\Lambda = \mathbb{R}^-$;
6. *Gamma Sub-Family.* If $p = 2$ with $\alpha = 0$, then the resulting ED model will be a gamma distribution model on support $\mathbb{R}^+$ with $\Omega = \mathbb{R}^+$ and $\Lambda = \mathbb{R}^+$;
7. *Positive Stable Sub-Family.* If $p \in (2, 3)$ with $\alpha \in \left(0, \frac{1}{2}\right)$, then the resulting ED model will be a positive stable model on support $\mathbb{R}^+$ with $\Omega = \mathbb{R}^+$ and $\Lambda = (-\infty, 0]$;
8. *Inverse Gaussian Sub-Family.* If $p = 3$ with $\alpha = \frac{1}{2}$, then the resulting ED model will be an IG model on support $\mathbb{R}^+$ with $\Omega = \mathbb{R}^+$ and $\Lambda = (-\infty, 0]$;
9. *Positive Stable-Generated (TS) Sub-Family.* If $p > 3$ with $\alpha = \left(\frac{1}{2}, 1\right)$, then the resulting ED model will be on support $\mathbb{R}^+$ with $\Omega = \mathbb{R}^+$ and $\Lambda = (-\infty, 0]$.

Remark 6.7 (Closed-Form LDA Models from the Exponential Dispersion Family) *Hence, we see that many popular LDA models are present as special cases of the ED model family. For instance, given a fixed number of losses in a year, one may model each loss as i.i.d. from one of the ED models above and obtain under the additive model structure closure under convolution to obtain a new ED model. Therefore, to get the compound process representation, one simply utilizes the discrete counting distributions representation to obtain the annual loss. Hence, we see that many closed-form LDA models may be obtained for appropriate choices of severity model from the ED family, see discussions in Kokonendji* et al. *(2004, section 3). In addition, we also see that for special cases when* $p \in (1, 2)$*, we get directly a compound Poisson model.*

It is also worth noting that in the Tweedie class of ED models, the limit as $p \to \infty$ results in the case of the ED models with variance function being characterized by the exponential

relationship given by

$$\lim_{p\to\infty} V(\mu;p) \propto \lim_{p\to\infty} \mu^p = \exp(\beta\mu),\ \beta \neq 0, \tag{6.123}$$

see discussion in Kokonendji *et al.* (2004).

6.3.1.1 Closed-Form LDA Poisson-Tweedie Exponential Dispersion Models $p \in (0,1)$. As discussed in Peters *et al.* (2009b), there are several methods to present the characterization of Poisson-Tweedie models for OpRisk settings. In the following, we present several characterizations given in Definition 6.17 which is from the perspective of a compound process for the annual loss of an LDA model. In Definition 6.17, we utilize the parametrization in Peters *et al.* (2009b) for the resulting Tweedie model with two simple notation changes: (i) the dispersion parameter ϕ is the inverse of λ in the previous ED models as opposed to σ^2; and (ii) in addition, the parameter α is replaced with γ.

Definition 6.17 (Tweedie Closed-Form LDA Models: Compound Poisson-Gamma $p \in (1,2)$) *Consider an LDA model for the annual loss with frequency distribution for the number of annual losses $N \sim Poisson(\lambda)$ and severity distribution for the ith loss random variable $X_i \sim F_X(x) = Gamma(k_i, \beta)$. These independent loss severity random variables have mean $\tau_i > 0$ with shape parameters $k_i > 0$ and scale parameter $\beta > 0$. Then the annual loss random variable Z takes a compound Poisson distribution*

$$Z = \mathbb{I}_{\{N>0\}} \sum_{i=1}^{N} X_i \sim F_Z(z;\lambda,\{k_i\},\beta) = \sum_{n=1}^{N} \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}\, Gamma\left(z;\ \sum_{i=1}^{n} k_i, \beta\right), \tag{6.124}$$

which will have a Poisson-Tweedie distribution with a jump at the origin $Z = 0$ given by $\mathbb{P}\mathrm{r}[N = 0] = \exp(-\lambda)$. The annual loss distribution for Z can be reparameterized in such a way that it takes a form of the ED family. In OpRisk, typically, the losses in a year are i.i.d., so we restrict to $k_i = k_j$ for all $i, j \in \{1, 2, \ldots, N\}$. The annual loss random variable Z has a probability weight at $N = 0$ given by

$$\mathbb{P}\mathrm{r}[Z = 0] = \mathbb{P}\mathrm{r}[N = 0] = \exp\{-\phi^{-1}\kappa(\theta;p)\}, \tag{6.125}$$

and for $Z > 0$, the random variable Z has continuous density

$$f_Z(z;\theta,\phi,p) = c(z;\phi,p)\exp\left\{\frac{z\theta - \kappa(\theta,p)}{\phi}\right\}. \tag{6.126}$$

Here $\theta < 0$, $\phi > 0$, and the normalizing constant is given by

$$c(z;\phi,p) = \sum_{r\geq 1}\left(\frac{(1/\phi)^{\gamma+1} z^{\gamma}}{(p-1)^{\gamma}(2-p)}\right)^r \frac{1}{r!\,\Gamma(r\gamma)z} \tag{6.127}$$

and the cumulant-generating function $\kappa(\cdot;p)$ is given by

$$\kappa(\theta;p) \stackrel{def.}{=} \frac{1}{2-p}[(1-p)\theta]^{\gamma}, \tag{6.128}$$

where $p \in (1,2)$ and $\gamma = (2-p)/(1-p)$. The parameters, in terms of the compound process representations, are given by

$$p = p(\gamma) = \frac{\gamma+2}{\gamma+1} \in (1,2), \tag{6.129}$$

$$\phi = \frac{\lambda^{1-p}\tau_i^{2-p}}{2-p} > 0, \tag{6.130}$$

$$\theta = \left(\frac{1}{1-p}\right)\mu^{(1-p)} < 0, \tag{6.131}$$

$$\mu = \lambda\tau > 0. \tag{6.132}$$

Then the mean and variance of Z are given by

$$\mathbb{E}[Z] = \frac{\partial}{\partial\theta}\kappa_p(\theta) = \kappa_p'(\theta) = [(1-p)\theta]^{1/(1-p)} = \mu, \tag{6.133}$$

$$\mathrm{Var}(Z) = \phi\kappa_p''(\theta) = \phi\,\mu^p. \tag{6.134}$$

That is, the annual loss Z has the mean μ, dispersion ϕ and variance function with the variance parameter p. Utilizing the above definitions, the distribution of Z can be rewritten in terms of μ, p and ϕ as

$$\mathbb{P}\mathrm{r}[Z=0] = \mathbb{P}\mathrm{r}[N=0] = \exp\left\{-\phi^{-1}\frac{\mu^{2-p}}{2-p}\right\}, \tag{6.135}$$

and for $z > 0$,

$$f_Z(z;\theta,\phi,p) = c(z;\phi,p)\exp\left\{\phi^{-1}\left[z\quad\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\right]\right\}. \tag{6.136}$$

From this definition, we may observe the following properties of the limiting cases as p approaches the boundaries of $(1,2)$ from within this domain. The extreme cases $p \to 1$ and $p \to 2$ correspond to the overdispersed Poisson and the Gamma models, respectively. Hence, in this spirit, Tweedie's compound Poisson model with $p \in (1,2)$ closes the gap between the Poisson and the gamma models. Often in practice, p is assumed to be known and fixed by the modeller.

Remark 6.8 (Parameter Estimation of the Closed-Form LDA Tweedie Models) *It will typically be wise to consider estimation of the parameters of the Tweedie model with $p \in (1,2)$ via the individual parameter estimation of the severity and frequency distributions separately. This will mean that one will have significantly more data available to the parameter estimation of the shape and scale parameters in the gamma severity model and the mean intensity of the frequency distribution. As opposed to estimation of the compound process model parameters given by the Tweedie model directly, which would also require evaluation of the normalization constant.*

Evaluation of the Normalizing Constant in Tweedie Models $p \in (0,1)$. Consider the normalizing constant of the density in the LDA model for the Poisson-Gamma model which we

have seen is a special case of the Tweedie ED models. The normalizing constant can be shown to be given by the function

$$c(z;\phi,p)=\sum_{r\geq 1}\left(\frac{(1/\phi)^{\gamma+1}y^{\gamma}}{(p-1)^{\gamma}(2-p)}\right)^{r}\frac{1}{r!\Gamma(r\gamma)y}=\frac{1}{y}\sum_{r\geq 1}W_r, \tag{6.137}$$

where $\gamma=\gamma(p)=(2-p)/(1-p)$ with the condition that $c(0;\phi,p)=1$ for $Z=0$.

The difficulty in the evaluation of the LDA annual loss distribution then lies in the evaluation of the normalizing constant $c(z;\phi,p)$ which contains an infinite sum. In Tweedie (1984), the summation was identified as Wright's (Wright 1935) generalized Bessel function, which cannot be expressed in terms of more common Bessel functions. To evaluate this summation, we follow the approach of Peters *et al.* (2009b) and Dunn & Smyth (2005) which directly sums the infinite series, including only terms which significantly contribute to the summation. Consider the term

$$\ln W_r=r\ln z-\ln\Gamma(1+r)-\ln\Gamma(\gamma r),$$

where

$$z=\frac{(1/\phi)^{\gamma+1}y^{\gamma}}{(p-1)^{\gamma}(2-p)}.$$

Replacing the gamma functions using Stirling's approximation and approximating γr by $\gamma r+1$, we get

$$\ln W_r\approx r\{\ln z+(1+\gamma)-\gamma\ln\gamma-(1+\gamma)\ln r\}-\ln(2\pi)-\frac{1}{2}\ln\gamma-\ln r,$$

which is also a reasonable approximation for small r. Treating r as continuous and taking the partial derivative w.r.t. r gives

$$\frac{\partial\ln W_r}{\partial r}\approx\ln z-\ln r-\gamma\ln(\gamma r).$$

Hence, the sequence W_r is unimodal in r. Solving $\partial W_r/\partial r=0$, to find (approximately) the maximum of W_r, results in the approximate maximum lying close to

$$R_0=R_0(\phi,p)=\frac{y^{2-p}}{(2-p)\phi}. \tag{6.138}$$

This gives a surprisingly accurate approximation to the true maximum of W_r, $r\in\mathbb{N}$. Finally, the aim is to find $R_L<R_0<R_U$ such that the following approximation is sufficiently accurate for the use in the evaluation of the likelihood terms,

$$c(y;\phi,p)\approx\widetilde{c}(y;\phi,p)=\frac{1}{y}\sum_{r=R_L}^{R_U}W_r. \tag{6.139}$$

The fact that $\partial\ln W_r/\partial r$ is monotonic and decreasing implies that $\ln W_r$ is strictly convex in r, and hence, the terms in W_r decay at a faster rate than geometric on either side of R_0. Dunn & Smyth (2005) derive the following bounds

$$c(y;\phi,p)-\widetilde{c}(y;\phi,p)<W_{R_L-1}\frac{1-q_L^{R_L-1}}{1-q_L}+W_{R_U+1}\frac{1}{1-q_U} \tag{6.140}$$

with

$$q_L = \exp\left(\frac{\partial \ln W_r}{\partial r}\right)\Big|_{r=R_L-1}, \; q_U = \exp\left(\frac{\partial \ln W_r}{\partial r}\right)\Big|_{r=R_U+1}. \tag{6.141}$$

These bounds are typically too conservative because the decay is much faster than geometric. In practice, an adaptive approach balancing accuracy and efficiency is to continue adding terms either side of the maximum until the lower and upper terms satisfy the double precision constraints $W_{R_L} \leq e^{-37} W_{R_0}$ (or $R_L = 1$) and $W_{R_U} \leq e^{-37} W_{R_0}$. When evaluating the summation for $\widetilde{c}(y;\phi,p)$, it was important to utilize the following identity to perform the summation in the log scale to avoid numerical overflow problems,

$$\ln \widetilde{c}(y;\phi,p) = -\ln y + \ln W_{R_0} + \ln\left(\sum_{r=R_L}^{R_U} \exp\left(\ln W_R - \ln W_{R_0}\right)\right).$$

One may make an additional observation when analyzing this model. As p approaches 1 (i.e. when the distribution approaches the overdispersed Poisson model), the likelihood may become multimodal. Therefore, to avoid numerical complications in actual calculations, one should consider restricting to $p \geqslant 1.1$. At the other extreme, when $p = 2$, the number of terms required to evaluate $c(y;\phi,p)$ may become very large; hence, to manage the computation burden, one may restrict $p \leqslant 1.95$.

6.3.1.2 Closed-Form LDA Poisson-Tweedie Exponential Dispersion Models $p \in (0,1)$.

In this section, one is interested primarily in the consideration of the severity model in the LDA structure that is in the Tweedie sub-family of the ED models. It was shown in Jørgensen (1986) that one can characterize the sub-family of severity models in the ED models that are ID sub-members by considering those that have positive support $X_i \sim ED(\lambda,\theta,p)$ such that each loss random variable $X_i \in [0,\infty)$ and the variance function is a power function of the mean with parameter p. Then the class of such models that are ID must satisfy the condition on the cumulant function used to characterize the family of ED models given in Theorem 6.7, see discussion in Jørgensen (1986, theorem 7.1).

Theorem 6.7 (Conditions for Infinite Divisibility in Exponential Dispersion Models) *Given a loss random variable $X_i \sim ED(\lambda,\theta,p)$ with support on $[0,\infty)$ and characterized by a variance function which is a power function of the mean given by $V(\mu) \propto \mu^p$. Then the random variable X_i is ID iff the following condition is satisfied for its cumulant function*

$$\frac{\partial^r}{\partial \theta^r}\kappa(\theta,p) \geq 0, \quad \forall r = 1,2,\ldots, \quad \textit{and} \;\; \theta \in int\Theta, \tag{6.142}$$

where $int\Theta$ represents the interior of the domain of parameter θ.

If we restrict ourselves to such models, we can obtain the generic closed-form LDA model in OpRisk settings under the ED model family in Proposition 6.16.

Proposition 6.16 (Closed-Form LDA Models in OpRisk ED Families) *Consider an LDA model for the annual loss distribution given by independent losses X_i (typically i.i.d.)*

$$Z = \sum_{i=1}^{N} X_i \sim F_Z(z;\zeta,\theta,\lambda,p) \tag{6.143}$$

for frequency distribution $N \sim F(\zeta)$ *with parameter* ζ *and severity distribution* $X_i \sim Tweedie(\theta, \lambda, p; A)$. *Defining the partial sum of the loss random variables*

$$S_n = \sum_{i=1}^{n} X_i \sim F_{S_n}, \tag{6.144}$$

one can write the annual loss distribution according to

$$F_Z(z; \zeta, \theta, \lambda, p) = \mathbb{P}\mathrm{r}[N = 0; \zeta]\mathbb{I}[Z = 0] + \left\{ \sum_{i=1}^{n} \mathbb{P}\mathrm{r}[N = n; \zeta] F_{S_n}(z; \theta, \lambda, p) \right\} \mathbb{I}[Z > 0] \tag{6.145}$$

and the annual loss density by

$$f_Z(z; \zeta, \theta, \lambda, p) = \mathbb{P}\mathrm{r}[N = 0; \zeta]\delta_{Z=0}(dz) + \sum_{i=1}^{n} \mathbb{P}\mathrm{r}[N = n; \zeta] f_{S_n}(z; \theta, \lambda, p) \tag{6.146}$$

with partial sums given by distribution and density functions for the corresponding member sub-family of the ED model according to

$$F_{S_n}(z; \theta, \lambda, p) = ED\left(z; \theta, \sum_{i=1}^{n} \lambda_i, p\right). \tag{6.147}$$

Several special case examples are now possible to consider as shown in the closed-form LDA models in Example 6.9.

EXAMPLE 6.9 Closed-Form Poisson-Inverse Gaussian LDA Models

Consider an LDA model for the annual loss distribution given by independent losses X_i (typically i.i.d.)

$$Z = \sum_{i=1}^{N} X_i \sim F_Z(z; \zeta, \theta, \lambda, p) \tag{6.148}$$

for frequency distribution $N \sim F(\zeta) = Poisson(\zeta)$ and severity distribution $X_i \sim Tw(\theta, \lambda, p; A) = InverseGaussian(x; \theta, \lambda, p = 3)$, where the severity distribution has distribution and density given by mean μ and scale λ according to

$$F_{X_i}(x; \mu, \lambda) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)$$

$$f_{X_i}(x; \mu, \lambda) = \left[\frac{\lambda}{2\pi x^3}\right]^{\frac{1}{2}} \exp\left(\frac{-\lambda(x - \mu)^2}{2\mu^2 x}\right), \quad x > 0. \tag{6.149}$$

Defining the partial sum of the loss random variables

$$S_n = \sum_{i=1}^{n} X_i \sim F_{S_n}, \tag{6.150}$$

one can write the annual loss distribution according to

$$F_Z(z;\zeta,\theta,\lambda,p) = \mathbb{P}\mathrm{r}[N=0;\zeta]\mathbb{I}[Z=0] + \left\{\sum_{i=1}^{n}\mathbb{P}\mathrm{r}[N=n;\zeta]F_{S_n}(z;\theta,\lambda,p)\right\}\mathbb{I}[Z>0] \tag{6.151}$$

and the annual loss density by

$$f_Z(z;\zeta,\theta,\lambda,p) = \mathbb{P}\mathrm{r}[N=0;\zeta]\delta_{Z=0}(dz) + \sum_{i=1}^{n}\mathbb{P}\mathrm{r}[N=n;\zeta]f_{S_n}(z;\theta,\lambda,p) \tag{6.152}$$

with partial sums given by distribution and density functions for the corresponding member sub-family of the ED model according to

$$\begin{aligned} F_{S_n}(z;\theta,\lambda_n,p) &= ED\left(z;\theta,\sum_{i=1}^{n}\lambda_i,p\right) \\ &= \Phi\left(\sqrt{\frac{\lambda_n}{z}}\left(\frac{z}{\mu_n}-1\right)\right) \\ &\quad + \exp\left(\frac{2\lambda_n}{\mu_n}\right)\Phi\left(-\sqrt{\frac{\lambda_n}{z}}\left(\frac{z}{\mu_n}+1\right)\right) \\ f_{S_n}(z;\theta,\lambda,p) &= \left[\frac{\lambda_n}{2\pi z^3}\right]^{\frac{1}{2}}\exp\left(\frac{-\lambda_n(z-\mu_n)^2}{2\mu_n^2 z}\right), \quad z>0, \end{aligned} \tag{6.153}$$

with $\mu_n = n\mu$ and $\lambda_n = n^2\lambda$ for each $n \in \mathbb{J}^+$. ■

In the next section, we present closed-form models for LDA structures that have sub-exponential severity models which allow for heavy-tailed features.

6.4 Sub-Exponential Families of LDA Models

In Chapter 3, the classes of sub-exponential models given by the α-stable and tempered α-stable models were introduced for severity modeling. In this section, we explore their development in closed-form LDA models, see discussions in Peters *et al.* (2011). These models are often considered because they exhibit extreme heavy tails with potential for infinite mean/variance. In Chernobai & Rachev (2004), a case study is performed for a range of different sub-exponential models for an undisclosed European bank's data on OpRisk losses between the period of 1950 and 2002 in which the loss data is classified simply into six categories: relationship, human, legal, processes, technology and external. Substantial amounts of loss data (for OpRisk settings) are utilized in these categories for the case study with a total of 1802 samples for the severity model losses. Simple Anderson–Darling and Kolmogorov–Smirnov (no weighting, see Cruz *et al.* 2015) were applied to test the fitted models for LogNormal, LogStable, Cauchy, stable and Pareto models. In several cases, there was statistically significant evidence from these

goodness-of-fit (GOF) tests to fail to reject the nominal claims that the data came from a heavy-tailed severity model such as a stable model. Therefore, in the following section, we illustrate how one may develop such heavy-tailed models under closed-form LDA structures utilizing the infinite divisibility and closure under convolution properties of such models. In addition, we note that such models have already been discussed earlier as part of the ED models with the index parameter $p > 3$.

In the following Section 6.4.1, we first discuss the practicality of working with such severity models which are closed under convolution but without restriction will take support on $\mathbb{R}$, whereas it is standard in risk modeling to consider monetary losses which are restricted to $\mathbb{R}^+$. Several scenarios are discussed in detail regarding solutions to this problem for the α-stable family. In addition, illustrations for the numerous choices for the counting frequency distribution mixture weights are presented for Poisson, binomial, negative-binomial, Poisson-gamma, Poisson-IG (Sitchel), Poisson-Tweedie and Hinde–Demetrio (HD) closed-form mixture weights are presented, see discussion in Kokonendji *et al.* (2004). Therefore, before presenting these closed-form LDA model examples, we first introduce some additional ideas regarding the severity distribution mixture weights for the different under- and overdispersion options for the frequency distribution that one may consider in OpRisk.

6.4.1 PROPERTIES OF DISCRETE EXPONENTIAL DISPERSION MODELS

In this section, we discuss the different choices of frequency distribution one can utilize in OpRisk settings when constructing closed-form LDA model, such as those covered in detail in Kokonendji *et al.* (2004). In this regard, we consider the following component, where for a given number of losses in a year n, one has the partial sums for the total annual loss given by

$$S_n = \sum_{i=1}^{n} X_n \sim F_{S_n}, \tag{6.154}$$

then in constructing the annual loss, we have seen that we must weight the possibility of each of the number of losses by the chance of having that many losses in the year, as quantified by the frequency distribution that provides the mixture weight $\mathbb{P}\mathrm{r}[N = n]$ in the LDA model to obtain the annual loss distribution given generically by

$$Z = \sum_{i=1}^{N} X_i \sim F_Z(z) = \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N = n] F_{S_n}(z). \tag{6.155}$$

The most popular frequency distribution choices we have seen so far are given by the mean–variance relationship matching or allowing for under- or overdispersion (mean number of losses greater or less than the variance) which gives simple mixture weights:

1. Poisson distribution with $N \sim Poisson(\lambda)$ such that one obtains for any $n \in \{0, 1, 2, \dots\}$

$$\mathbb{P}\mathrm{r}[N = n] = \exp(-\lambda)\frac{\lambda^n}{n!}. \tag{6.156}$$

2. Binomial distribution with $N \sim Binomial(p, K)$ such that one obtains for any $n \in \{0, 1, 2, \dots, K\}$

$$\mathbb{P}\mathrm{r}[N = n] = \frac{K!}{(K-n)!n!} p^n (1-p)^{K-n}. \tag{6.157}$$

3. Negative binomial distribution with $N \sim NegativeBinomial(p, r)$ such that one obtains for any $r \in \{0, 1, 2, \dots\}$

$$\mathbb{P}\mathrm{r}[N = n] = \frac{(n + r - 1)!}{(r - 1)!n!}(1 - p)^r p^n. \tag{6.158}$$

Then one may consider more flexible frequency distributions given by doubly stochastic choices such as the Poisson-Gamma and the Poisson-IG cases given by mixture weights:

- Poisson-Gamma distribution for $N \sim Poisson - Gamma(\lambda, \alpha, \beta)$ which is defined by a gamma intensity where we assume that the intensity parameter for the mean number of losses in a year is a random variable given by

$$\Lambda \sim \Gamma(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\beta\Lambda)\Lambda^{\alpha-1}, \tag{6.159}$$

where $\alpha > 0$ and $\beta > 0$, which results in the mixed Poisson model involving the probability of the event $\{N = n\}$ being given by

$$\begin{aligned}\mathbb{P}\mathrm{r}[N = n] &= \int_0^\infty \exp(-\lambda)\frac{\lambda^n}{n!}\frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\beta\lambda)\lambda^{\alpha-1} d\lambda \\ &= \frac{(\alpha + n - 1)!}{(\alpha - 1)!n!}\left(\frac{\beta}{1 + \beta}\right)^\alpha \left(\frac{1}{1 + \beta}\right)^n. \end{aligned} \tag{6.160}$$

See details in Peters *et al.* (2011, theorem 6).

- Poisson-GIG intensity for $N \sim Poisson - GIG(\lambda; \alpha, \beta, \gamma)$ which is defined by a GIG intensity function, that is, a generalized version of the Sitchel model where we assume that the intensity parameter for the mean number of losses in a year is a random variable given by

$$\begin{aligned}\Lambda \sim GIG(\lambda; \alpha, \beta, \gamma) &= \left(\frac{2\sqrt{(1 - \beta)}}{\alpha\beta}\right)^\gamma \left[2K_\gamma\left(\alpha\sqrt{(1 - \beta)}\right)\right]^{-1} \lambda^{\gamma-1} \\ &\quad \times \exp\left(-\left[\frac{1}{\beta} - 1\right]\lambda - \frac{\alpha^2\beta}{4\lambda}\right), \end{aligned} \tag{6.161}$$

where $\gamma \in \mathbb{R}$, $\alpha > 0$ and $\beta \in [0, 1]$, which results in the mixed Poisson model involving the probability of the event $\{N = n\}$ being given by

$$\begin{aligned}\mathbb{P}\mathrm{r}[N = n] &= \int_0^\infty \exp(-\lambda)\frac{\lambda^n}{n!}\left(\frac{2\sqrt{(1 - \beta)}}{\alpha\beta}\right)^\gamma [2K_\gamma(\alpha\sqrt{(1 - \beta)})]^{-1}\lambda^{\gamma-1} \\ &\quad \times \exp\left(-\left[\frac{1}{\beta} - 1\right]\lambda - \frac{\alpha^2\beta}{4\lambda}\right) d\lambda \\ &= \frac{(1 - \beta)^{\gamma/2}}{K_\gamma(\alpha\sqrt{(1 - \beta)})}\frac{(\alpha\beta)^n}{2^n n!}K_{n+\gamma}(\alpha). \end{aligned} \tag{6.162}$$

See details in Sichel (1982, equation 2.4).

Kokonendji *et al.* (2004), discuss two distinct and general families of discrete integer values distributions that are obtainable from the ED family discussed previously. These correspond to the Poisson-Tweedie family cases discussed earlier for Poisson, binomial, negative binomial, Poisson-Gamma and Poisson-IG models. One can also obtain from this class the mixed discrete ED family of Poisson-positive stable models. All these types of models are obtained by considering the following generic mixing structure given in Definition 6.18. Note that, in this definition, we replace the use of λ in referring to the class of ED models with r to avoid confusion with the intensity parameter for the Poisson distribution that we denote by λ and treat as stochastic below.

Definition 6.18 (Discrete-Integer-Valued Poisson-Tweedie Frequency Distributions) *The frequency distribution for the number of losses N in a year can be defined according to a Poisson-Tweedie mixture with $N \sim PT(\theta, r, p)$ if the counting distribution takes the following form:*

$$\Pr[N = n] = \int_0^\infty \exp(-\lambda)\frac{\lambda^n}{n!}\, Tweedie(\theta, r, p)d\lambda, \tag{6.163}$$

where $\lambda \sim Tweedie(\theta, r, p)$ is the class of Tweedie ED models which correspond to the selection of power variance functions given by

$$V(\mu) \propto \mu^p. \tag{6.164}$$

It was shown in Hougaard *et al.* (1997) that if one considers $N \sim PT(\theta, r, p)$, then for $\theta \in \Theta_p$ defined by

$$\Theta_p = \begin{cases} \mathbb{R}, & p = 0, 1, \\ [0, \infty), & p < 0 \ \ or \ \ p \in (0, 1), \\ (-\infty, 0), & p \in (1, 2] \ \ or \ \ p \to \infty, \\ (-\infty, 0], & p \in (2, \infty), \end{cases} \tag{6.165}$$

and $r > 0$ for a value of $p \geq 1$, then one has the following properties for the frequency distribution:

1. Aggregating the number of losses across k different periods or across k-different loss processes, etc. when they are assumed to be independent and following the Poisson-Tweedie distribution $N_i \sim PT(\theta, r_i, p)$ with the above parameter restrictions will result in the joint distribution being distributed as

$$N_T = N_1 + N_2 + \cdots + N_k \sim PT\left(\theta, \sum_{i=1}^{k} r_i, p\right), \tag{6.166}$$

   which shows that such frequency distributions for $N \sim PT(\theta, r, p)$ will be ID.
2. The distribution for $N \sim PT(\theta, r, p)$ can be shown to be unimodal for $p \geq 2$. This has relevant properties for evaluation and truncation of such infinite mixture weights in evaluation of closed-form LDA models.

3. In Kokonendji *et al.* (2004), the Laplace transform (moment-generating function (MGF)) of $N \sim PT(\theta, r, p)$ is shown to satisfy the three cases:

$$\mathbb{E}[\exp(sN)] = \begin{cases} \exp\left(\dfrac{r}{2-p}\{[(1-p)(\exp(s)-1+\theta)]^{\alpha} - [(1-p)\theta]^{\alpha}\}\right), & p \neq 1, 2, \\ \dfrac{\theta^r}{(\exp(s)-1+\theta)^r}, & p = 2, \\ \exp(r[\exp(s-1+\theta) - \exp(\theta)]), & p = 1, \end{cases} \tag{6.167}$$

where α is selected to satisfy $(p-1)(1-\alpha) = 1$. Knowledge of this transform allows one to differentiate and evaluate at the origin to obtain the moments of the frequency distribution to detect the presence of under- or overdispersion for different sets of input parameters θ, r and p. In the case that $p = 1$, one obtains the Neyman type A distribution; for $p = 2$, the negative binomial distribution arises; and for $p = 3$, one obtains the Sitchel distribution.

Remark 6.9 *The closure under convolution of the frequency distribution above is an important general result which is very important for combining loss data from different loss processes together into one model in order to obtain, for instance, the annual loss across a financial institution or across a business unit or division. This property is of course true for Poisson distributions in which the sum of k-independent Poisson-distributed random variables each given by $N_i \sim Poisson(\lambda_i)$ produces*

$$N_T = N_1 + N_2 + \cdots + N_k \sim Poisson(\lambda_1 + \lambda_2 + \cdots + \lambda_k). \tag{6.168}$$

However, the above result generalizes this concept for a significantly larger class of frequency distributions with numerous over- and underdispersion features possible.

One may also consider a distinct class of models which also take discrete support and are in the family of ED models for consideration when modeling the frequency distribution; these are referred to in Kokonendji *et al.* (2004) as the HD class as detailed in Definition 6.19. This class of models can be considered as an overdispersed alternative family to the negative binomial which is a special case when $p = 2$ and the arcsine distribution also arises as a special case when $p = 3$.

Definition 6.19 (Hinde–Demetrio Class of Frequency Distributions in the ED Family) *The frequency distribution for the number of losses N in a year can be defined according to a HD sub-family of discretely supported ED models with $N \sim HD(\theta, r, p)$ if the counting distribution involves an ED family with $p > 1$ and variance function characterized by the relationship*

$$V(\mu) = \mu + \mu^p. \tag{6.169}$$

The resulting HD family of frequency distribution takes the form of an NEF generated by a tilting measure $\nu(dx)$ according to the discrete-integer-valued representation

$$\mathbb{P}\mathrm{r}[N \in \mathscr{A}] = \sum_{x \in \mathscr{A}} \exp(\theta x - \lambda\kappa(\theta))\nu(dx; p), \tag{6.170}$$

with cumulant function given by

$$\kappa(\theta) = \frac{1}{\lambda} \ln \int \exp(\theta x) \nu(dx; p). \tag{6.171}$$

From the cumulant function choice, one can find the relationship between the canonical parameter θ and the mean and variance functions μ and $V(\mu)$ given in terms of the mapping τ given by

$$\mu = \tau(\theta) = \frac{d}{d\theta}\kappa(\theta),$$
$$V(\mu) = \frac{d}{d\theta}\tau(\tau^{-1}(\mu)) = \mu + \mu^p. \tag{6.172}$$

In the HD family, the choice of tilting measure is given by

$$\nu(dx; p) = \sum_{k=0}^{\infty} \frac{a(a+1)\cdots(a+k-1)}{k!} \frac{1}{1+k(p-1)} \delta_{1+k(p-1)}(dx), \tag{6.173}$$

with $a = (p-1)^{-1}$.

It was shown in Kokonendji *et al.* (2004) that for all $r > 0$, one obtains the measure

$$\nu(dx; p, r) = \exp[r\nu(dx; p)] = \sum_{n=0}^{\infty} \frac{r^n}{n!} \nu^{(n)*}(dx; p), \tag{6.174}$$

which can be used to develop additional NEF models as above with variance functions given by

$$V(m) = m + r^{1-p} m^p, \tag{6.175}$$

which will have a distribution which is concentrated on the additive semi-group on $\mathbb{N} + p\mathbb{N}$.

In the remaining sections of this chapter, we will consider different choices for such frequency distribution models in the construction of flexible closed-form LDA models for heavy-tailed cases.

6.4.2 CLOSED-FORM LDA MODELS FOR LARGE LOSS NUMBER PROCESSES

In OpRisk, there are several settings in which the number of losses that may occur in a given year can be substantial, even if each individual loss amount is not too large. Examples might include in practice credit card frauds, small loans frauds, computer errors and employee frauds. In such settings, one can often make a reasonable model based on the family of LDA models comprised a geometric frequency distribution for the number of losses in the year with a severity distribution which could be assumed to be either finite variance or infinite variance. In each case, the resulting infinite limit for the number of losses as the mean of the geometric distribution for the frequency goes to infinity will result in a convergence to one of the three families of closed-form annual loss distributions:

1. Asymmetric Laplace distribution (ASD) family (finite variance severity models);
2. Geometric stable (GS) models (infinite variance severity models);
3. Generalized ν-sable distributions (non-geometric compound process random sums).

The class of GS distributions were discussed briefly in Cruz *et al.* (2015, chapter 15, 16) in the context of insurance and also in Kozubowski & Podgorski (1999) and Mittnik & Rachev (1991). To clarify, we present the closed-form LDA model generic structure for the large loss numbers modeling settings in Proposition 6.17. The results considered here are utilizing scale mixture representations of stable random variables with a specially selected sequence of normalizing and centring constants to ensure the convergence of such compound processes to either the ALD or the GS distribution as given by the geometric compounds given in Equation 6.176, see discussion in Kozubowski & Podgórski (2001). In particular, one may consider the two model classes presented as suitable closed-form model approximations to compound processes with large numbers of annual losses.

6.4.2.1 Asymmetric Laplace Distributions for Geometric Random Sums of Finite Variance Losses. In general, one requires the following properties to hold N_p is independent of X_i random variables, the X_i random variables are independent copies of the annual loss distribution Z such that the asymptotic limit

$$Z_p = a_p \sum_{i=1}^{N_p} (X_i + b_p) \xrightarrow{d} Z \sim F_Z, \quad as \;\; p \downarrow 0, \tag{6.176}$$

holds for sequences of coefficients (functions of p) given by

$$a_p = Cp^{1/2}, \;\; b_p = \frac{\left(p^{1/2} - C\right)\mu}{C}, \;\; C = \sqrt{\frac{2}{2 + \left(\frac{\mu}{\sigma}\right)^2}}. \tag{6.177}$$

In addition, it can be shown that all the distributions in the domain of attraction of this family of geometric compound losses that converge to a non-degenerate distribution as $p \downarrow 0$ will also be asymptotically self-decomposable. It can be shown that all distributions in this domain of convergence asymptotically will also be in the L-Class of distributions, see Chapter 4 which ensures they are unimodal.

Proposition 6.17 (Light-Tailed Closed-Form Large Number of Losses LDA Model) *Consider an OpRisk loss process with an LDA structure for the annual loss Z in which the frequency distribution for the number of losses in the year is given by $N \sim Geometric(p)$ and the severity distribution is generically given by any distribution such that each loss is i.i.d. and given by $X_i \sim F(x)$. The only condition placed on the distribution of the severity model is that $\mathbb{V}\mathrm{ar}[X] < \infty$. The frequency distribution is given by*

$$\mathbb{P}\mathrm{r}[N = n] = p(1-p)^{n-1}, \;\; n = 1, 2, \ldots \tag{6.178}$$

In this case, the large loss number annual loss distribution is asymptotically approximated well by

$$\lim_{p \to 0} Z_p = \lim_{p \to 0} \sum_{i=1}^{N_p} X_i \xrightarrow{d} Z \sim F_Z(z; \sigma, \mu), \tag{6.179}$$

where the limiting annual loss random variable Z for the large loss number model has a characteristic function given by

$$\Phi_Z(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta Z)] = \frac{1}{1 + \sigma^2\theta - \mathrm{i}\mu\theta} \tag{6.180}$$

with the location parameter $\mu \in \mathbb{R}$ and scale parameter $\sigma > 0$. The distribution and density are given by

$$f_Z(z;\mu,\sigma) = \begin{cases} \frac{1}{\sigma}\frac{\kappa}{1+\kappa^2}\exp\left(-\frac{\kappa}{\sigma}z\right), & z \geq 0, \\ \frac{1}{\sigma}\frac{\kappa}{1+\kappa^2}\exp\left(\frac{1}{\kappa\sigma}z\right), & z < 0, \end{cases}$$

$$F_Z(z;\mu,\sigma) = \begin{cases} 1 - \frac{1}{1+\kappa^2}\exp\left(-\frac{\kappa}{\sigma}z\right), & z \geq 0, \\ \frac{1}{1+\kappa^2}\exp\left(\frac{1}{\kappa\sigma}z\right), & z < 0, \end{cases}$$

where $\kappa = \dfrac{2\sigma}{\mu + \sqrt{4\sigma^2 + \mu^2}}$.

The following special cases arise from this LDA model family:

1. If $\mu = 0$, one obtains the Laplace distribution.
2. If $\mu = 0$ and $\sigma = 0$, one obtains a degenerate dirac mass at the origin.
3. If $\mu > 0$ and $\sigma = 0$, then one obtains an exponential distribution with mean μ. In this case, in the geometric sum in the limit as $p \downarrow 0$ for

$$Z_p = a_p \sum_{i=1}^{N_p}(X_i + b_p) \xrightarrow{d} Z \sim F_Z, \quad as \;\; p \downarrow 0, \tag{6.181}$$

   the appropriate constants that arise are $a_p = p$ and $b_p = 0$.
4. If $\mu = 0$ and $\sigma > 0$, then one obtains the Laplace distribution with location at the origin and density

$$f_Z(z;\mu = 0,\sigma) = \frac{1}{2\sigma}\exp\left(-\left|\frac{z}{\sigma}\right|\right). \tag{6.182}$$

   In this case, in the geometric sum in the limit as $p \downarrow 0$ for

$$Z_p = a_p \sum_{i=1}^{N_p}(X_i + b_p) \xrightarrow{d} Z \sim F_Z, \quad as \;\; p \downarrow 0, \tag{6.183}$$

   the appropriate constants that arise are $a_p = \sqrt{p}$ and $b_p = 0$.

There are two relevant observations to make about this ALD model for practical settings, one may simulate annual loss severities from this model using the scale mixture representation where

$$Z \overset{d}{=} \mu E + \sqrt{2\sigma^2 E}G \tag{6.184}$$

for independent random variables $E \sim Exp(1)$ and $G \sim Normal(0,1)$. In addition, it is useful to know that parameter estimation can be done efficiently via the maximum-likelihood estimation (MLE) procedure where the parameter estimates are solved in closed form in Proposition 6.18, see Kozubowski & Podgorski (1999).

Proposition 6.18 (MLE Parameter Estimation of the ALD Annual Loss Model) *The MLE of the ALD model parameters from sample data is efficient and the parameter estimators are given in closed form with respect to a sample* $Z_1, Z_2, \ldots, Z_n$ *by*

$$\begin{aligned}\widehat{\mu}_n &= \frac{1}{n}\sum_{i=1}^{n} Z_i, \\ \widehat{\sigma}_n^2 &= \left(\frac{1}{n}\sum_{i=1}^{n} Z_i^+\right)^{\frac{1}{4}}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^-\right)^{\frac{1}{4}}\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} Z_i^+} + \sqrt{\frac{1}{n}\sum_{i=1}^{n} Z_i^-}\right)\end{aligned} \tag{6.185}$$

with $Z^+ = \max\{Z, 0\}$ *and* $Z^- = \max\{-Z, 0\}$.

Remark 6.10 *In practical settings, using the above MLE procedure on the annual losses may not be accurate as typically financial institutions only have a few years of data. However, this is the case where there is, however, very large sample sizes for the non-aggregated losses, so it may be more suitable to fit the severity model to the individual (not annual aggregate) data using MLE or other methods for an appropriate finite variance severity model. Then to make the geometric frequency approximation and to assess the quality of this assumption (asymptotic rate of convergence of the fitted severity model sums to the ALD-limiting distribution family), see the following discussions.*

Of course, in practice, one has a finite number of losses present so it would be important to assess under what circumstances one can motivate and utilize such an asymptotic large loss number approximation for a closed-form model in the OpRisk setting; there are numerous methods one could adopt to verify the validity of the ALD closed-form LDA model structure:

1. Fit the ALD model to the data via maximum likelihood and perform a GOF test (based on tests presented in Cruz *et al.* (2015, chapter 8)).
2. One could fit the LDA model for a desired severity and frequency model and then assess the fitted LDA models moments up to nth order using the fitted LDA model against those obtained from just fitting an ALD model via an MLE procedure. In this case, the MGF of the LDA model with a given severity and frequency distribution is given by the composite between the pgf of the frequency distribution for number of losses N and the MGF of the severity distribution for loss amounts X according to

$$\mathrm{MGF}(Z) := M_Z(\theta) = \mathbb{E}[\exp(\theta Z)] = M_N(\ln[M_X(\theta)]). \tag{6.186}$$

This could then be differentiated n-times and evaluated for each derivative at the origin to produce the moments when they exist, which will be in terms of the parameters of the given frequency and severity models selected for the LDA model. These moments can then be compared (or indeed the coefficients of the Taylor-expanded MGF for the LDA model) to those of the MLE-fitted ALD asymptotic closed-form LDA model characterized by nth integer moments given by

$$\mathbb{E}[Z^n] = n!\left(\frac{\sigma}{\kappa}\right)^n \frac{1 + (-1)^n \kappa^{2(n+1)}}{1 + \kappa^2}. \tag{6.187}$$

3. Alternatively, one could perform a heuristic assess the sample mean, variance, coefficient of variation and coefficient of skewness from the sample and compare to a fitted ALD model via MLE procedure to see if the moments are reasonable representative. The population ALD moments are given by

$$\text{Mean:} \quad \mathbb{E}[Z] = \mu,$$

$$\text{Mean deviation:} \quad \mathbb{E}[Z - \mathbb{E}[Z]] = \frac{2\sigma \exp(\kappa^2 - 1)}{\kappa(1 + \kappa^2)},$$

$$\text{Coefficient of variation:} \quad \frac{\sqrt{\mathbb{V}\text{ar}[Z]}}{|\mathbb{E}[Z]|} = \sqrt{2\frac{\sigma^2}{\mu^2} + 1},$$

$$\text{Coefficient of skewness:} \quad \frac{\mathbb{E}[Z - \mathbb{E}[Z]]^3}{(\mathbb{E}[Z - \mathbb{E}[Z]]^2)^{\frac{3}{2}}} = 2\frac{\kappa^{-3} - \kappa^3}{(\kappa^{-2} + \kappa^2)^{\frac{3}{2}}},$$

$$\text{Coefficient of kurtosis:} \quad \frac{\mathbb{E}[Z - \mathbb{E}[Z]]^4}{\mathbb{V}\text{ar}[Z]^2} = 6 - \frac{12}{(\kappa^{-2} + \kappa^2)^2}.$$

Remark 6.11 *Technically, the exact distribution would arise in the case in which the geometric sum in the limit as $p \downarrow 0$ for*

$$Z_p = a_p \sum_{i=1}^{N_p} (X_i + b_p) \xrightarrow{d} Z \sim F_Z, \quad as \quad p \downarrow 0, \tag{6.188}$$

where independent X_i loss random variables had the same distribution as the annual loss Z such as in the Laplace and exponential geometric sums covered by Arnold (1973) and Hu & Lin (2003).

6.4.2.2 Geometric Stable for Geometric Random Sums of Infinite Variance Losses. Next we consider the case in which the severity model is not finite variance but can have finite mean. In this case, the geometric compound process will result in an asymptotic limiting closed-form distributional family for the large number of OpRisk losses given by a GS distribution as discussed in Proposition 6.19.

Proposition 6.19 (Heavy-Tailed Closed-Form Large Number of Losses LDA Model) *Consider an OpRisk loss process with an LDA structure for the annual loss Z in which the frequency distribution for the number of losses in the year is given by $N \sim Geometric(p)$ and the severity distribution is generically given by any distribution such that each loss is i.i.d. and given by $X_i \sim F(x)$. The only condition placed on the distribution of the severity model is that $\mathbb{V}\text{ar}[X]$ is not finite. The frequency distribution is given by*

$$\mathbb{P}\text{r}[N = n] = p(1 - p)^{n-1}, \quad n = 1, 2, \ldots \tag{6.189}$$

In this case, the large loss number annual loss distribution is asymptotically well approximated by

$$\lim_{p \to 0} Z_p = \lim_{p \to 0} \sum_{i=1}^{N_p} X_i \xrightarrow{d} Z \sim F_Z(z; \alpha, \beta, \gamma, \delta), \tag{6.190}$$

where the limiting annual loss random variable Z for the large loss number model has a characteristic function given by

$$\Phi_Z(\theta) = \mathbb{E}[\exp(\mathrm{i}\theta Z)] = \frac{1}{1 + \gamma^\alpha |\theta|^\alpha \psi(\theta; \alpha, \beta) - \mathrm{i}\delta\theta} \tag{6.191}$$

with

$$\psi(\theta; \alpha, \beta) = \begin{cases} 1 - \mathrm{i}\beta sgn(\theta) \tan\left(\frac{\pi\alpha}{2}\right), & if\ \alpha \neq 1, \\ 1 + \mathrm{i}\beta \frac{\pi}{2} sgn(\theta) \ln|\theta|, & if\ \alpha = 1, \end{cases} \tag{6.192}$$

where $\alpha \in (0, 2]$, is the index of stability as in the α-stable family. If $\alpha = 2$, one obtains the ALD family previously with finite variance. The parameter $\beta \in (-1, 1)$ is the skewness parameter, $\gamma > 0$ is the scale parameter and $\delta \in \mathbb{R}$ is the location parameter.

It can be shown that the GS laws have the following special cases:

1. if $\beta = 0$, $\delta = 0$, then one obtains the Linnik family of distributions;
2. if $\beta = 1$ and if either $\alpha = 1$ and $\gamma = 0$ or $\alpha \in (0, 1)$ and $\delta = 0$, then one obtains the exponential distribution.

To understand the relationship between the geometric stable and the standard α-stable model, it is best to observe that the GS annual loss random variable can be re-expressed by the scale mixture representation with respect to a standard α-stable random variable $Z \sim S(\alpha, \beta, \gamma, \delta)$ with $\gamma = 1$ and $\delta = 0$ and an independent exponential random variable $E \sim Exp(1)$ according to the equality in distribution:

$$Z \overset{d}{=} \begin{cases} \delta Z + Z^{1/\alpha} \gamma X, & \alpha \neq 1, \\ \delta Z + Z\gamma X + \gamma Z \beta \frac{2}{\pi} \ln(Z\gamma), & \alpha = 1. \end{cases} \tag{6.193}$$

This will also allow for direct simulation of such annual loss random variables through exact simulation of a standard α-stable random variable followed by simulation independently of an exponential random variable which are combined as above for a draw from a GS annual loss distribution.

The relationship between the characteristic function of the GS and the Paretian (standard) α-stable (PS) model is given in Equation 6.194, see discussion in Mittnik & Rachev (1991).

$$\Phi_Z(\theta; GS) = \frac{1}{1 - \ln\phi_Z(\theta; PS)}. \tag{6.194}$$

6.4.2.3 ν-Stable for Random Sums of Infinite Variance Losses. If one finds that the geometric frequency distribution assumption is not necessarily suitable, one can also define alternative classes of convergent sequences of the form above with different sequences of coefficients a_p and b_p which are generically known as ν-stable laws, see discussions in Kozubowski & Podgorski (1999). This class of distributions generalizes the ALD and GS distributions discussed earlier, where the frequency distribution is given by a positive, integer-valued family of random variable $N_p \sim \nu(p)$ with distributions characterized by parameter $p \in (0, 1)$ such that as $p \to 0$ one has that $N_p \to \infty$ in probability. In addition, for the appropriate closed-form asymptotic LDA models to be obtained the convergence of $pN_p \xrightarrow{d} Z$ as $p \to 0$ with $Z \sim F_Z$ must apply, which allows one to define the class of ν-stable laws.

Proposition 6.20 (Closed-Form Large Number of Losses LDA Model: ν-Stable Family) *Consider an OpRisk loss process with an LDA structure for the annual loss Z in which the frequency distribution for the number of losses in the year is given by $N \sim \nu(p)$ and the severity distribution is generically given by any distribution such that each loss is i.i.d. and given by $X_i \sim F(x)$. As $p \to 0$, if one has that $\nu(p) \to \infty$ in probability and $p\nu(p) \xrightarrow{d} \nu$ with distribution ν having Laplace transform of the form $[1+x]^{-1}$, then the annual loss distribution is said to be ν-stable $Z \sim \nu S(\alpha, \beta, \gamma, \mu)$ and given by*

$$\lim_{p\to 0} Z_p = \lim_{p\to 0} \sum_{i=1}^{N_p} X_i \xrightarrow{d} Z \sim F_Z, \tag{6.195}$$

where Z has characteristic function for the ν-stable model denoted by $\Phi_Z(\theta; \nu)$ which is given with respect to a PS characteristic function $\Phi_Z(\theta; PS)$ according to

$$\Phi_Z(\theta; \nu S) = \lambda(-\ln \Phi_Z(\theta; PS)) \tag{6.196}$$

with

$$\lambda(z) = \int_0^\infty \exp(-zx) dF_Z(x) \tag{6.197}$$

with

$$\Phi_Z(\theta; PS) = \exp(-\gamma^\alpha |\theta|^\alpha \psi(\theta; \alpha, \beta) + \mathrm{i}\delta\theta) \tag{6.198}$$

and

$$\psi(\theta; \alpha, \beta) = \begin{cases} 1 - \mathrm{i}\beta \tan\left(\frac{\pi\alpha}{2}\right) sgn(\theta), & \alpha \neq 1, \\ 1 + \mathrm{i}\beta \frac{2}{\pi} \ln|\theta| sgn(\theta), & \alpha = 1, \end{cases} \tag{6.199}$$

with $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\gamma \geq 0$ and $\mu \in \mathbb{R}$.

One can also show that the ν-stable family of LDA models is scale invariant as shown in Proposition 6.21.

Proposition 6.21 *Consider a LDA model in which the number of losses is sufficiently large that one can approximate asymptotically the annual loss LDA model by $Z \sim \nu S(\alpha, \beta, \gamma, \delta)$, then under scaling of the annual loss distribution, one obtains*

$$cZ \sim \begin{cases} \nu(\alpha,\ sgn(c)\beta, |c|\gamma, c\delta), & \alpha \neq 1, \\ \nu(\alpha,\ sgn(c)\beta, |c|\gamma, c\delta -), & \alpha \neq 1. \end{cases} \tag{6.200}$$

In Kozubowski & Panorska (1996), they show that one can derive the following tail features of the resulting ν-stable distribution as detailed in Proposition 6.22.

Proposition 6.22 (Tail Properties of ν-Stable LDA Models) *Consider an annual loss distribution given by $Z \sim \nu S(\alpha, \beta, \gamma, \delta)$ with $\delta = 0$ and $\alpha \neq 1$, or $\beta = 0$ and $\alpha = 1$. Then*

consider $\alpha \in (0,2)$ *and the following tail limits are finite for the* ν*-stable annual loss distribution, if* $\eta = \int_0^\infty z dF_Z(z) < \infty$*, giving*

$$\begin{aligned}\lim_{z\to\infty} z^\alpha \overline{F}_Z(z) &= C_\alpha \frac{1+\beta}{2}\gamma^\alpha \eta < \infty,\\ \lim_{z\to\infty} z^\alpha F_Z(-z) &= C_\alpha \frac{1-\beta}{2}\gamma^\alpha \eta < \infty,\end{aligned} \tag{6.201}$$

where

$$C_\alpha = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos\left(\frac{\pi\alpha}{2}\right)}, & \alpha \neq 1,\\ \dfrac{2}{\pi}, & \alpha = 1. \end{cases} \tag{6.202}$$

The implications of these tail properties are of practical significance to OpRisk LDA models as they show that if these tail limits apply and if, in addition, one can show that $\mathbb{E}[|Z\ln Z|] < \infty$ when $\alpha = 1$ or $\mathbb{E}[Z^{1\vee\alpha}] < \infty$ when $\alpha \neq 1$, then the following absolute moments of the annual loss distribution are finite:

$$\mathbb{E}[|Z|^p] = C\gamma^p \mathbb{E}\left[Z^{\frac{p}{\alpha}}\right] < \infty, \quad for \;\; p \in (0,\alpha). \tag{6.203}$$

with

$$C = \frac{2^{p-1}\Gamma\left(1-\frac{p}{\alpha}\right)}{p\int_0^\infty u^{-p-1}\sin^2 u du}\left(1+\beta^2\tan^2\frac{\alpha\pi}{2}\right)^{p/2\alpha}\cos\left[\frac{p}{\alpha}\arctan\left(\beta\tan\frac{\alpha\pi}{2}\right)\right]. \tag{6.204}$$

This is the ν-stable analogue of the fractional lower order moment (FLOM) defined for the α-stable models in Cruz *et al.* (2015, chapter 8), see discussions in Kozubowski & Panorska (1996). It can be shown that when Z is GS, then in this case, one has $\mathbb{E}[Z^{p/\alpha}] = \Gamma\left(\frac{p}{\alpha}+1\right)$.

6.4.3 CLOSED-FORM LDA MODELS FOR THE α-STABLE SEVERITY FAMILY

Results demonstrating how closed-form analytic expressions for the annual loss distribution of the LDA model given by the Poisson-stable compound process are obtained. We can then utilize properties of the α-stable and tempered α-stable severity models given in (Cruz *et al.* 2015, chapter 8) to derive closed-form analytic expressions for the distribution and density of the annual loss distribution.

As discussed in detail in Chapter 4, the α-stable models can be considered as generalizations of the Gaussian distribution, α-stable models are defined as the class of location-scale distributions which are closed under convolutions. In general, it is widely known that it will suffice to consider the properties of the α-stable severity distribution with respect to tail index parameter α and β, and in general, when studying the properties of this family of severity models, this can be achieved without the loss of generality, see discussion (Cruz *et al.* 2015, chapter 8). However, in OpRisk, we generally wish to restrict to the class of α-stable models that take positive real support or to transform the losses such that such a restriction is not required, because this will ensure that we only work with a non-negative loss processes. In standard parametric models which take support on the entire real line and are tractable, this

is achieved trivially by considering a left tail truncation at the origin, such that the severity distribution is given by Equation 6.205.

$$f_X(x|X > 0) = \frac{f_X(x)}{1 - F_X(0)}, \tag{6.205}$$

where we can get the evaluation pointwise of the density $f_X(x)$ and the distribution $F_X(x)$ using the results in Lemma 4.7 and Corollary 6.1, see derivations in (Zolotarev, 1986, remark 2.2.27, p. 78). As a consequence of these representations of the distribution and density for the α-stable severity models, one can also show that they admit the following practically useful results in Corollary 6.1. These results are important to aid in understanding the effect of truncation of the severity distribution so that resulting loss processes may have a positive support.

Corollary 6.1 (Closed-Form α-Stable Severity Density and Distribution at Origin) *w.l.o.g. the density function of an α-stable severity distribution (standardized such that $\gamma = 1$ and $\delta = 0$) can be evaluated at the origin utilizing*

$$f_X(0; \alpha, \beta, 1, 0; B) = \frac{1}{\pi}\gamma\left(1 + \frac{1}{\alpha}\right)\cos\left(\frac{\pi}{2}\beta\frac{K(\alpha)}{\alpha}\right),\ \alpha \neq 1. \tag{6.206}$$

Secondly, of direct application in the context of OpRisk is to observe the result of (Zolotarev 1983, equation 2.2.30) that shows that

$$F_X(0; \alpha, \beta, 1, 0; B) = \mathbb{P}\mathrm{r}[X < 0] = \frac{1}{2}\left(1 - \beta\frac{K(\alpha)}{\alpha}\right),\ \alpha \neq 1. \tag{6.207}$$

However, in the context of the class of α-stable models, such a truncated representation would no longer admit closure under convolution, hence, precluding the properties required for analytic representations of the compound process model. To understand this, consider the addition of two left-truncated Gaussian-distributed random variables will not in general remain Gaussian distributed.

Therefore, we consider alternatives for modeling described in the following scenarios which will still provide severity models with suitable support, either with exactly positive support or controllably small chances of negative loss that still admit closure under convolution and, therefore, analytic expressions for annual loss and risk measures. This can be achieved in several ways:

1. *Scenario 1.* Consider modeling the loss values via a stable severity model on the log scale in which case one can consider the entire range of α-stable models. This may be appropriate for cases in which the original severity distribution was believe to be, for example, reasonably modeled by distributions such as a LogNormal (exactly stable—Normal—on log scale), log-Cauchy hyper-heavy-tailed family of models (exactly stable on log scale), log-logistic, generalized log-logistic (Burr Type XII), Champernowne and Dagum distributions or others of such forms.
2. *Scenario 2.* As a second alternative, one may believe that the stable model is better suited to modeling on the original data loss scale, in which case one can enforce a positive support for the α-stable severity distribution with $S0$ parametrization discussed in Chapter 4 through restrictions on the severity model parameters in which the tail index $\alpha <$

1; the skewness parameter satisfies $\beta = 1$ and the location and scale parameters satisfy $\delta = \gamma \tan \frac{\pi\alpha}{2}$. Note that considering ranges for $\alpha < 1$ is appropriate only for situations in which one believes that modeling of infinite mean loss processes is required.

3. *Scenario 3.* As a third alternative, for situations in which one believes that either a finite mean model is required with $\alpha > 1$ or the restrictions on the severity parameters above are not suitable for the application. In such settings, if one considers for the severity model a given tail index α and location parameter δ and a reasonable restriction for severity models in OpRisk that satisfy $\beta \geq 0$, then one can prove that the majority of the mass lies to the left of the mode of the severity distribution

$$\mathbb{P}\mathrm{r}[X > m(\alpha, \beta, \gamma, \delta)] > \mathbb{P}\mathrm{r}[X < m(\alpha, \beta, \gamma, \delta)], \tag{6.208}$$

see discussion in Chapter 4. In addition, one also knows that in practice, in OpRisk, the mode $m(\alpha, \beta, \gamma, \delta) >> 0$ can lie very far from the origin, so that if one models the loss data on the original scale, then one can approximately quantify the left tail probability for $x \in [-\infty, -\varepsilon]$ for ε in the neighbourhood of 0 by

$$\mathbb{P}\mathrm{r}[X < -\varepsilon] \sim \gamma^{\alpha} c_{\alpha} (1 - \beta) \varepsilon^{-\alpha}. \tag{6.209}$$

with $c_{\alpha} = \sin\left(\frac{\pi\alpha}{2}\right) \Gamma(\alpha)/\pi$. In particular, note that when one considers almost perfectly skewed models, this probability is infinitesimally small as $\beta \uparrow 1$ for cases in which the neighbourhood of the origin is an extreme for the severity process. Finally, to completely specify the probability mass associated to the left tail as quantified by $X \in (-\infty, 0)$, one can utilize the result in Corollary 6.1 which shows that w.l.o.g. one has

$$\mathbb{P}\mathrm{r}[X < 0] = \frac{1}{2}\left(1 - \beta \frac{K(\alpha)}{\alpha}\right), \ \alpha \neq 1. \tag{6.210}$$

4. *Scenario 4.* A fourth alternative is to consider α-stable models for the severity which are obtained under Scenario 3. Then if modeling is only focussed on the annual loss distribution, one can truncate the annual loss distribution to a positive support and this will still produce analytic results for the truncated annual loss distribution, although the individual severity model is not on their own truncated to a positive support. This will be advantageous in certain situations as discussed.
5. *Scenario 5.* A fifth alternative is to consider asymmetric TS models, with a strong tempering on the left tail, which will also be discussed later.

These approaches provide alternatives to truncation of the severity distribution at the origin, which would result in a loss of the closure under convolution properties of this model that give rise to the analytic representations presented in this chapter for the annual loss process. Even if the third alternative is adopted, this is not problematic in practice for modeling in OpRisk as mass in the negative support can be made arbitrarily small as discussed. To illustrate some of these scenarios discussed, we provide an illustration in Example 6.10.

EXAMPLE 6.10 α-Stable Severity Models for OpRisk (Scenario 1)

Consider the sub-exponential family of severity models characterized by the α-stable severity model for loss data on the log scale with asymmetry parameter

$\beta = 0$. For example, if the original severity distribution was exactly LogNormal, then the resulting log-transformed loss distribution is Normal (a special case of the symmetric α-Stable model). In the cases in which the severity distribution is approximately LogNormal, perhaps with heavier tails or more asymmetry, then the resulting symmetric stable approximation may be still a suitable model. Figure 6.1a demonstrates the original severity distribution for a LogNormal severity with location parameter $\mu = 0$ and scale parameter $\sigma = 1$, then Figure 6.1c shows the log-transformed symmetric stable approximation, in this case, exactly a Normal distribution. If, however, the original data is approximated well by any of the following forms: log-logistic, etc. in form, then the log-transformed data is still well approximated by a symmetric α-stable model as depicted in Figure 6.1b and d. Figure 6.1b depicts an example of a heavy-tailed 'LogNormal-like' severity distribution (i.e. a Burr Type XII), whilst Figure 6.1d depicts an example of the log-transformed data kernel density estimate versus the estimated symmetric α-stable model.

■

If the loss data is not appropriately modeled under **Scenario 1** and instead one wishes to consider **Scenario 3**, then Example 6.11 illustrates the left tail probability of the severity distribution of an α-stable distribution $\mathbb{P}\mathrm{r}[X < 0]$ for a range of severity distributions.

EXAMPLE 6.11 α-Stable Severity Models for OpRisk (Scenario 3)

Consider the sub-exponential family of severity models characterized by the α-stable severity model for loss data with generic scale parameter $\gamma = 1$ and with asymmetry parameter $\beta \in [0, 1)$ and location parameter $\delta > 0$. For illustration, we consider two sub-cases, when an infinite mean model is appropriate for very heavy-tailed loss processes $\alpha < 1$ and when an infinite mean model is not appropriate. Considering illustrative examples with $\alpha \in \{0.5, 1.5\}$ and $\beta \in \{0.25, 0.5, 0.95\}$ for a range of location parameters $\delta \in \{0, 1, 2, 3, 4, 5\}$, we evaluate the tail probability and compare this to the asymptotic result discussed. First, we present for $\delta = 5$ the severity distributions for the different (α, β) combinations in Figure 6.2. In both the cases, we see that after scaling the loss data to provide unit scale parameter $\gamma = 1$, the resulting severity distributions very quickly result in negligible left tail probabilities $\mathbb{P}\mathrm{r}[X < 0; \alpha, \beta, 1, \delta; 0]$ as $\beta \uparrow 1$ and/or δ grows large. We quantify this formally in Figure 6.3 where we plot the cumulative distribution function (cdf) for a range of parameter settings for the example of $\alpha = 1.5$. Finally, we conclude this example by considering a worst-case type scenario in which the severity distribution location parameter is set such that $\delta = 0$ and a scale of $\gamma = 1$ is considered and we plot the probability of a negative loss versus $\alpha \neq 1$ and $\beta \geq 0$. First, we observe that for $\alpha < 1$, the trajectory of the truncation probability $\mathbb{P}\mathrm{r}[X < 0]$ is identical for all $\alpha < 1$ as a function of β and as expected starts at 0.5 for the zero-centred symmetric α-stable ($\beta = 0$) and shrinks non-linearly to 0 for β in the neighborhood of 1, i.e. $\beta \in n.e.(1)$. The behaviour of the truncation probability is depicted in Figure 6.4

■

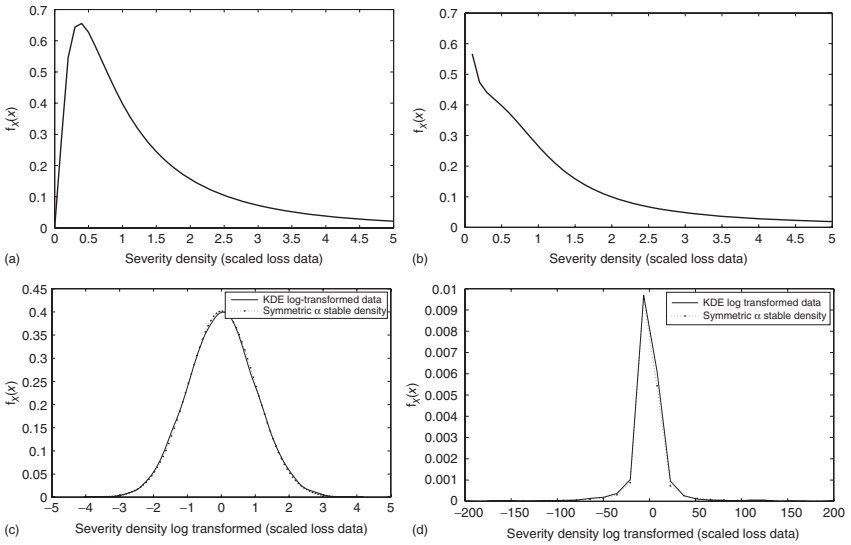

FIGURE 6.1 Scenario 1 examples of log-transformed distributions versus symmetric stable approximations.

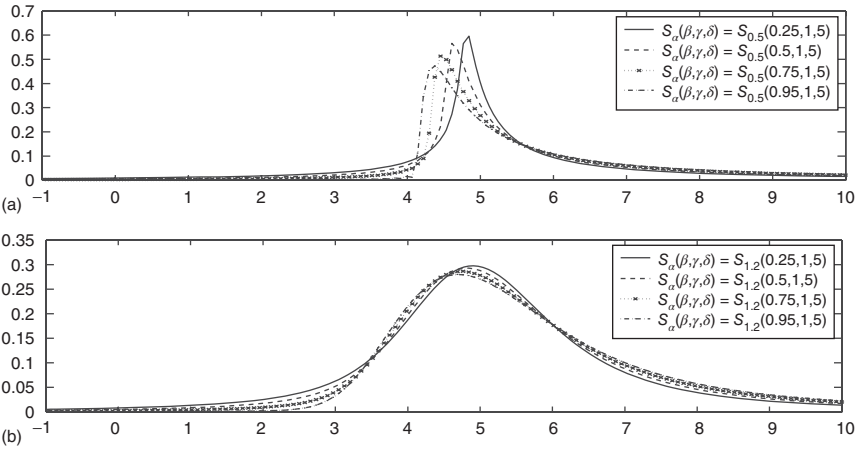

FIGURE 6.2 **Scenario 3 examples of α-stable severity distributions for location $\delta = 5$. (a) Infinite mean models with $\alpha = 0.5$ and a range of skewness parameters $\beta \in \{0.25, 0.5, 0.95\}$. (b) Finite mean heavy-tailed models with $\alpha = 1.2$ and a range of skewness parameters $\beta \in \{0.25, 0.5, 0.95\}$.**

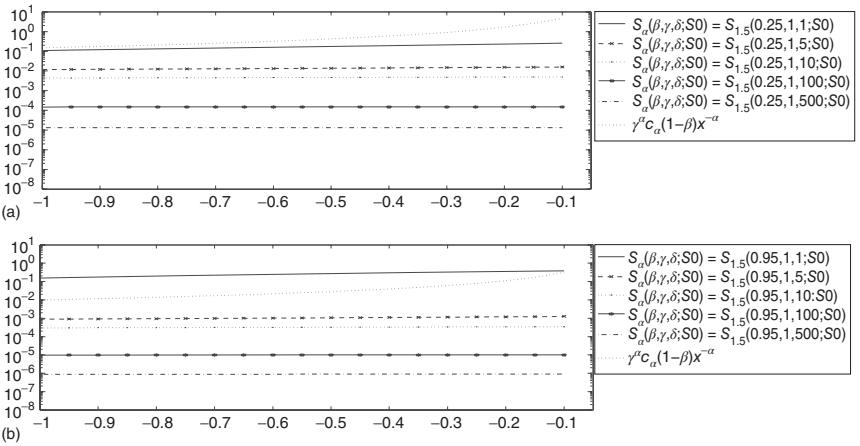

FIGURE 6.3 **Scenario 3 examples of α-stable severity distribution left tail cdf (log scale) for finite mean model varying skewness β and location δ. (a) Finite mean models with $\alpha = 1.5$ and $\beta = 0.25$ (close to symmetric) and a range of location parameters versus asymptotic left tail approximation. (b) Finite mean models with $\alpha = 1.5$ and $\beta = 0.95$ (close to symmetric) and a range of location parameters versus asymptotic left tail approximation.**

So bearing in mind these scenarios that may be adopted as modeling approaches, one can now proceed to present a general representation of analytic solutions to OpRisk LDA models constructed from the family of α-stable models.

We assume the ith loss of the jth risk process in year t is a random variable with α-stable distribution, denoted by $X_i^{(j)}(t) \sim S_\alpha(x; \beta, \gamma, \delta, S0)$, where $S_\alpha(x; \beta, \gamma, \delta, S0)$ denotes the univariate four-parameter stable distribution family under parametrization $S0$, see Section 4.4.2 for details. Given the properties of the α-stable model, one can derive an analytic exact Poisson mixture representation of the annual loss process under α-stable severity models with

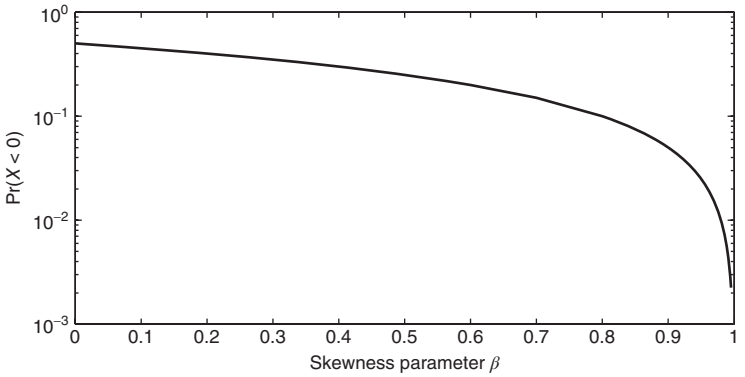

FIGURE 6.4 **Scenario 3 examples of α-stable severity distribution left tail cdf (log scale) for infinite mean models varying skewness β.**

the required positive support as provided in Theorem 6.8. In addition, we will provide an analytic expression for the tail distribution of these models and the properties of the ES in special cases of this LDA model.

We note that the following analytic results for the LDA Poisson-stable model is a consequence of the results derived in Zolotarev (1986, property 2.3) as detailed in Proposition 6.23 and later restated in the $S0$ and $S1$ parameterizations in Nolan (2015).

Proposition 6.23 *Consider stable random loss variables $X_i \sim S_\alpha(\beta_i, \gamma_i, \delta_i; A)$ and real numbers b and a_i for $i \in \{1, 2, \ldots, n\}$. These can be used to construct the stable distribution from the following linear combination with stable distribution given by*

$$Z_n = \sum_{i=1}^{n} a_i X_i + b \sim S_\alpha(\widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; A) \tag{6.211}$$

with parameters given by

$$\begin{aligned}
\widetilde{\gamma}_n &= \sum_{i=1}^{n} \gamma_i |a_i|^\alpha, \\
\widetilde{\gamma}_n \widetilde{\beta}_n &= \sum_{i=1}^{n} \gamma_i \beta_i |a_i|^\alpha sgn(a_i) \\
\widetilde{\gamma}_n \widetilde{\delta}_n &= \begin{cases} \sum_{i=1}^{n} \gamma_i \delta_i a_i + b, & \text{if } \alpha \neq 1, \\ \sum_{i=1}^{n} \gamma_i \delta_i a_i + b - \frac{2}{\pi} \sum_{j=1}^{n} \gamma_j \beta_j a_j \ln |a_j|, & \text{if } \alpha = 1. \end{cases}
\end{aligned} \tag{6.212}$$

We also observe that one can further generalize this result via the result in Zolotarev (1986, p.64) as detailed in Proposition 6.24

Proposition 6.24 *Considering the results stated for the A-type parametrization in Proposition 6.23, if one now considers generalizing this result to the case in which the weights in the linear combination a_i are random variables, one can still obtain a stable distributional result. This is given in the simplest form according to the following relationship. Consider $X_i \sim S_\alpha(0, \gamma, 0; A)$ and*

random variables A_i for $i \in \{1, 2\}$ that must only satisfy the condition $\mathbb{P}\mathrm{r}[A_1 + A_2 = 0] = 0$. Furthermore, assume $X_i \perp X_j \perp A_i, \ \forall i, j$, then one has the following equality in distribution:

$$\left(\frac{A_1}{A_1 + A_2}\right)^{1/\alpha} X_1 - \left(\frac{A_2}{A_1 + A_2}\right)^{1/\alpha} X_2 \stackrel{d}{=} X_1. \tag{6.213}$$

From these two results, one can then utilize these results under the $S0$ parametrization to define an analytic LDA model when working with α-stable severity models.

Theorem 6.8 (Analytic LDA Annual Loss Poisson-α-stable) *Given the number of losses is distributed $N \sim Poisson(\lambda)$ and for all $i \in \{1, \ldots, N\}$, each loss random variable is distributed as α-stable. Furthermore, assume that the support of the severity distribution is specified according to one of the following scenarios:*

1. *$X_i \sim S_\alpha(x; \beta_i, \gamma_i, \delta_i; S0)$ with restriction that $\beta_i = 1$, $\alpha < 1$ and $\delta_i = \gamma_i \tan \frac{\pi\alpha}{2}$ s.t. for all $i \in \{1, \ldots, N\}$, one has $\mathbb{P}\mathrm{r}[X_i < 0] = 0$ due to the duality theorem Zolotarev (1983);*
2. *the log-transformed severity in which $\widetilde{X}_i = \ln(X_i) \sim S_\alpha(x; \beta_i, \gamma_i, \delta_i; S0)$ with no restrictions on the parameters $\alpha, \beta, \gamma, \delta$ is modeled;*
3. *a controllably small probability of negative loss is considered such that $X_i \sim S_\alpha(x; \beta_i, \gamma_i, \delta_i; S0)$ with no restrictions on the parameters α, γ, and with mild loss distributional restrictions on $\delta > 0$ and skewness parameter $\beta \in (0, 1]$. Noting that for any β, a corresponding region for $\delta \in (a, \infty)$ can be found such that the probability $\mathbb{P}\mathrm{r}[X_i < 0] \to 0$ as $\beta \uparrow 1$ and typically for OpRisk in the neighbourhood of 1, that is, $\beta \in n.e.(1)$.*

Given these considerations above for OpRisk LDA modeling and the severity distribution support, we state the following general result which includes these cases. For α-stable parameters $\alpha \in [0, 2]$, $\beta \in [-1, 1]$, $\gamma > 0$ and $\delta \in \mathbb{R}$, given α-stable-distributed random variables with common α (dropping the tilde notation for the log transform), the resulting distribution of the compound process (annual loss under an LDA framework) is given by

$$Z = \sum_{i=1}^{N} X_i \sim \sum_{n=1}^{\infty} \exp(-\lambda) \tfrac{\lambda^n}{n!} S_\alpha(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S0), \tag{6.214}$$

where the parameters of each mixture component are analytically known as expressions of the base severity model according to

$$\begin{aligned} \widetilde{\gamma}^\alpha &= \sum_{i=1}^{N} |\gamma_i|^\alpha, \ \widetilde{\beta} = \frac{\sum_{i=1}^{N} \beta_i |\gamma_i|^\alpha}{\sum_{i=1}^{N} |\gamma_i|^\alpha}, \\ \widetilde{\delta} &= \begin{cases} \sum_{i=1}^{N} \delta_i + \tan\frac{\pi\alpha}{2} \left(\widetilde{\beta}\widetilde{\gamma} - \sum_{i=1}^{N} \beta_j \gamma_j\right) & \text{if } \alpha \neq 1, \\ \sum_{i=1}^{N} \delta_i + \frac{2}{\pi} \left(\widetilde{\beta}\widetilde{\gamma} \ln \widetilde{\gamma} - \sum_{i=1}^{N} \beta_j \gamma_j \ln |\gamma_i|\right) & \text{if } \alpha = 1. \end{cases} \end{aligned} \tag{6.215}$$

Hence, one observes that as a result of closure under convolution of the α-stable severity model, the resulting distribution for the annual loss can be presented exactly as a mixture representation.

Corollary 6.2 *The results in Theorem 6.8 are trivially extended to binomial and negative binomial compound process LDA models, with appropriate adjustments to the mixture weights for these frequency distributions. This is illustrated below for a special sub-family of models.*

Remark 6.12 *Generally, for the majority of values of* $\alpha \in [0, 2]$*, the pointwise evaluation of the severity distribution and, therefore, the pointwise evaluation of the annual loss distribution are not analytic. However, for univariate settings, there are efficient numerical results that will admit evaluation, see Chapter 4 for details. For example, one can evaluate the severity distribution pointwise utilizing the integral representations or series representations in (Zolotarev 1983, equations 2.2.1a and 2.2.1b)*

Remark 6.13 *Note that in many practical settings, one would consider for extreme consequence loss processes, the possibility of a finite number of total losses that could occur in a given year. This would impose a truncation on the counting process for the number of losses in a given year, such that* $N \in \{1, 2, \ldots, N_{\max}\}$*. This would simply result in an adjustment to the mixture weights according to those presented in Corollary 6.3.*

The following result for truncated counting process for the losses can also be applied to all infinite analytic mixture models discussed in this Chapter. It will often be both practically sensible to consider such a truncation as well as practically convenient in exact evaluation of densities for these models.

Corollary 6.3 (Analytic LDA Annual Loss Truncated-Poisson-α-stable) *Given the number of losses is distributed* $N \sim Poisson(\lambda)\mathbb{I}[N \leq N_{\max}]$ *and for all* $i \in \{1, \ldots, N\}$ *each loss random variable is distributed as* α*-stable, then the resulting distribution of the compound process (annual loss under an LDA framework) is given by*

$$Z = \sum_{i=1}^{N} X_i \sim \sum_{n=1}^{N_{\max}} \frac{N_{\max}!}{\Gamma(N_{\max} + 1, \lambda)} \exp(-\lambda) \frac{\lambda^n}{n!} S_\alpha(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S0), \tag{6.216}$$

where $\Gamma(N_{\max} + 1, \lambda)$ *is the incomplete gamma function.*

In addition, we note that if modeling under Scenario 4, one can obtain an annual loss distribution with strictly positive support for any α-stable loss distribution assumptions on the parameters utilizing the results in Theorem 6.8 and noting that by truncating each mixture component, the resulting distributional support will be strictly positive, see Corollary 6.9.

Theorem 6.9 (Analytic LDA Annual Loss (Positive Support) Poisson-α-Stable) *Given the number of losses is distributed* $N \sim Poisson(\lambda)$ *and for all* $i \in \{1, \ldots, N\}$ *each loss random variable is distributed according* $X_i \sim S_\alpha(x; \beta_i, \gamma_i, \delta_i; S0)$ *for unrestricted parameters, the resulting distribution of the annual loss under an LDA compound process framework with positive support,* $Z \in [0, \infty)$*, is given by*

$$Z = \sum_{i=1}^{N} X_i \sim \sum_{n=1}^{\infty} \exp(-\lambda) \frac{\lambda^n}{n!} c(\alpha, \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n, \varepsilon) S_\alpha(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S0)$$
$$\times \mathbb{I}[z \geq \varepsilon \in \text{n.e.}(0) > 0], \tag{6.217}$$

where the parameters are as defined in Theorem 6.8 and the normalization constants for each component are given by noting the Duality theorem results for α-stable distributions in Zolotarev (1983, section 2.4, remark 2) which show that one can get this probability as a series expansion up to any desired precision according to the expression $x = \varepsilon \in n.e.(0) > 0$,

$$
\begin{aligned}
&c(\alpha, \beta, \gamma, \delta, \varepsilon) \\
&= \begin{cases}
\left(-\frac{1}{\pi\alpha} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\Gamma(n\alpha+1)}{n\Gamma(n+1)} \sin(\pi n\rho) \left(\frac{\varepsilon-\delta}{\gamma^{\alpha}} \right)^{-\alpha n} \right)^{-1}, & \text{if } \alpha < 1, \\
\left(\frac{1}{2} \left(1 + \beta \frac{K(\alpha)}{\alpha} \right) \right. & \\
\left. + \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{\Gamma\left(\frac{n}{\alpha}+1\right)}{n\Gamma(n+1)} \sin(\pi n\rho) \left(\frac{\varepsilon-\delta}{\gamma^{\alpha}} \right)^{n} \right)^{-1}, & \text{if } \alpha > 1, \\
\left(-\frac{1}{\pi} b_0 + \frac{1}{\pi} \sum_{n=1}^{\infty} (-1)^{n-1} b_n \left(\frac{\varepsilon-\delta}{\gamma} \right)^{n} \right)^{-1}, & \alpha = 1, \\
& \beta > 0.
\end{cases}
\end{aligned}
$$

Note this series expansion for the normalization constant for the truncation of the annual loss process is obtained through rearranging the expression in (Zolotarev 1983, equation 2.4.3, p. 89). Typically, one can truncate this series expansion by progressively adding terms until the change in the evaluated cdf is within a desired tolerance or threshold.

Remark 6.14 *In several OpRisk settings, in practice, it is sensible to also consider an upper truncation, that is, there may exist a total accumulated loss in a given year that would be possible to experience for a given loss process. This may, for example, be associated with OpRisk arising from infrastructure failure and IT failures. These upper bounds on the total loss can arise, for example, from the total actual cost of the asset or alternatively can be considered in more abstract scenarios in which the loss is limited in liability either due to legal contractual obligations or through insurance mitigation. In all such cases, it is of importance to also consider specification of the LDA models for the α-stable families presented with upper tail truncation on the annual loss process.*

Theorem 6.10 (Analytic LDA Annual Loss (Limited Annual Liability) Poisson-α-Stable) *Given the number of losses is distributed $N \sim Poisson(\lambda)$ and for all $i \in \{1, \ldots, N\}$ each loss random variable is distributed according $X_i \sim S_{\alpha}(x; \beta_i, \gamma_i, \delta_i; S0)$ for unrestricted parameters, the resulting distribution of the annual loss under an LDA compound process framework with positive support in an annual limited range, $Z \in [0, z_{\max})$, is given by*

$$
Z = \sum_{i=1}^{N} X_i \sim \sum_{n=1}^{\infty} \exp(-\lambda) \frac{\lambda^n}{n!} c_2(\alpha, \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n, z_{\max}) S_{\alpha}(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S0) \mathbb{I}[z \in [\varepsilon, z_{\max}]]
$$

for $\varepsilon \in \text{n.e.}(0) > 0$ and parameters are as defined in Theorem 6.8. In this case, the normalization constants for each component are given by

$$c_2(\alpha,\beta,\gamma,\delta,z)$$
$$= \begin{cases} \left(\frac{1}{\pi\alpha}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(n\alpha+1)}{n\Gamma(n+1)}\sin(\pi n\rho)\left[\left(\frac{\varepsilon-\delta}{\gamma^{\alpha}}\right)^{-\alpha n}-\left(\frac{z-\delta}{\gamma^{\alpha}}\right)^{-\alpha n}\right]\right)^{-1}, & \text{if } \alpha<1, \\ \left(\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma\left(\frac{n}{\alpha}+1\right)}{n\Gamma(n+1)}\sin(\pi n\rho)\left[\left(\frac{z-\delta}{\gamma^{\alpha}}\right)^{n}-\left(\frac{\varepsilon-\delta}{\gamma^{\alpha}}\right)^{n}\right]\right)^{-1}, & \text{if } \alpha>1, \\ \left(\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}b_n\left[\left(\frac{z-\delta}{\gamma}\right)^{n}-\left(\frac{\varepsilon-\delta}{\gamma}\right)^{n}\right]\right)^{-1}, & \alpha=1, \\ & \beta>0. \end{cases}$$

We complete the analytic representations of the family of LDA models constructed from α-stable severity models by considering the results one can obtain for Poisson, binomial, negative binomial, and doubly stochastic Poisson and negative Binomial LDA α-stable models as developed in Peters *et al.* (2010a). In particular, we will illustrate such models in the special sub-family of α-stable Lévy family of severity models. Note that all the following models may be extended to general α-stable settings as discussed earlier.

We begin by considering discrete time loss processes in which the total number of losses is assessed in fixed intervals of time, such as quarterly or monthly time increments. In such cases, we may wish to consider non-independent counts per month and so one can consider the annual loss distribution modeled by a frequency distribution with binomial distribution or with negative binomial distribution, depending on whether we believe the variance in the number of losses exceeds the mean number of losses or is less than the mean number of losses. In the continuous arrival processes setting, we consider annual losses to arrive according to independent increments of time according to an exponential arrival processes, resulting in a Poisson frequency distribution. Clearly, under a Poisson-based model, we are implicitly assuming that the mean number of losses per year is equal to the variance in the number of annual losses, when the rate parameter represents the average number of annual losses. We will then extend this Poisson family of models to the doubly stochastic family of frequency distributions and derive closed-form expressions for the annual loss distribution in such cases.

Starting with the Poisson-Lévy LDA model, one notes that this particular sub-family of models, presented in Peters *et al.* (2011), will produce an analytic result as detailed in the following theorems and illustrated in the following examples.

Lemma 6.10 (Lévy Sub-family of α-Stable Severity Models) *Given loss random variables which are distributed according to the Lévy α-stable sub-family of models, $X \sim S(0.5,1,\gamma,\delta;S0)$ with positive real support $x \in [\delta,\infty]$. The density and distribution functions are analytic and given, respectively, for $\delta < x < \infty$, by*

$$f_X(x) = \sqrt{\frac{\gamma}{2\pi}}\frac{1}{(x-\delta)^{3/2}}\exp\left(-\frac{\gamma}{2(x-\delta)}\right),$$
$$F_X(x) = \operatorname{erfc}\left(\sqrt{\frac{\gamma}{2(x-\delta)}}\right).$$

The median is given by $\widetilde{\mu} = \frac{\widetilde{\gamma}_n}{2}(\operatorname{erfc}^{-1}(0.5))^2$ and the mode $M = \frac{\widetilde{\gamma}_n}{3}$, where $\operatorname{erfc}(x) = 1-\operatorname{erf}(x) = 1-\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$.

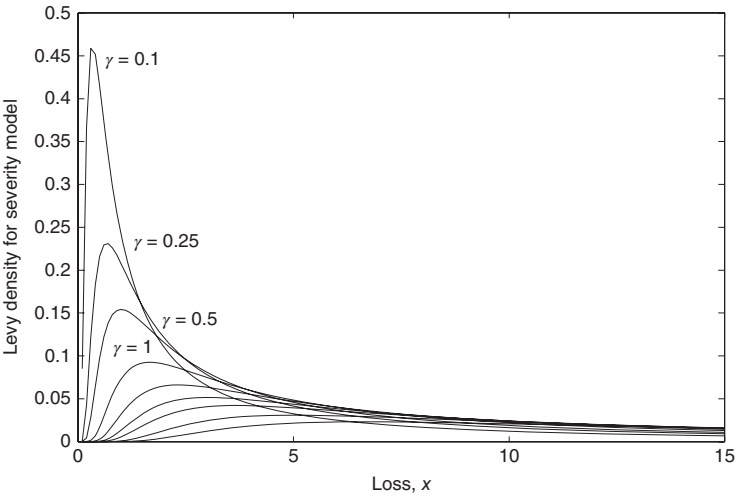

FIGURE 6.5 **Example of the Lévy severity model as a function of scale parameter, with location $\delta = 0$.**

This result follows from Nolan (2015, chapter 1, p. 5). As a consequence, one may derive the analytic representation for the family of annual loss processes given by Theorem 6.11. Furthermore, an illustration of the properties of this sub-family of sub-exponential severity model is obtained by plotting several realizations of the severity distribution for a range of scale parameters in Figure 6.5.

Theorem 6.11 (LDA Annual Loss Poisson-Lévy) *The distribution of the annual loss process Z represented by a compound process model with LDA structure in which the frequency is $N \sim Poisson(\lambda)$ and the severity model $X_i \sim S(0.5, 1, \gamma, \delta; S0)$, then the exact density of the annual loss process can be expressed analytically as a mixture density comprised α-stable components with Poisson mixing weights for $N_t^{(j)} > 0$,*

$$f_Z(z) = \sum_{n=1}^{\infty} \exp(-\lambda) \frac{\lambda^n}{n!} \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z - \widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)} \right) \right] \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

with $\widetilde{\beta}_n = 1$ and

$$\begin{aligned} \widetilde{\gamma}_n^{0.5} &= \sum_{i=1}^{n} |\gamma_i|^{0.5} = n|\gamma|^{0.5}, \\ \widetilde{\delta}_n &= \sum_{i=1}^{n} \delta_i + \tan\frac{\pi}{4}\left(\widetilde{\gamma}_n - \sum_{j=1}^{n} \gamma_j \right) \\ &= n\delta + \tan\frac{\pi}{4}(n^2|\gamma| - n\gamma), \end{aligned}$$

and $f_Z(0) = \mathbb{P}\mathrm{r}(N_t^{(j)} = 0) = \exp(-\lambda)$ for $N = 0$. The exact form of the annual loss cdf is also expressible in closed form,

$$F_Z(z) = \sum_{n=1}^{\infty} \exp(-\lambda)\frac{\lambda^n}{n!}\text{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}}\right)\mathbb{I}[\widetilde{\delta}_n < z < \infty] + \exp(-\lambda)\mathbb{I}[z=0].$$

Next in Theorem 6.12, we derive the analogous results in the setting in which the intensity process of the arrival of the losses is inhomogeneous in time as may be the case in many settings in which regulatory standards change, management structures change or management policy changes over time.

Theorem 6.12 (LDA Annual Loss Doubly-Stochastic-Poisson-Lévy) *The distribution of the annual loss processes Z represented by a doubly stochastic compound processes model with LDA structure in which the frequency is $N(t) \sim Poisson(\lambda)$ and $\lambda \sim Gamma(\alpha, \beta)$ and the severity model $X_i(t) \sim S(0.5, 1, \gamma, \delta; 0)$, then the exact density of the annual loss processes can be expressed analytically as a mixture density comprised α-stable components with Poisson mixing weights for $N_t > 0$,*

$$\begin{aligned} f_Z(z) &= \sum_{n=1}^{\infty}\left[\int_0^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\exp\left(-\beta\lambda\right)\lambda^{\alpha-1}d\lambda\right] \\ &\quad\times\left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}}\frac{1}{\left(z-\widetilde{\delta}_n\right)^{3/2}}\exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}\right)\right]\times\mathbb{I}[\widetilde{\delta}_n < z < \infty] \\ &= \sum_{n=1}^{\infty}\frac{(\alpha+n-1)!}{(\alpha-1)!n!}\left(\frac{\beta}{1+\beta}\right)^{\alpha}\left(\frac{1}{1+\beta}\right)^{n} \\ &\quad\times\left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}}\frac{1}{\left(z-\widetilde{\delta}_n\right)^{3/2}}\exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}\right)\right]\times\mathbb{I}[\widetilde{\delta}_n < z < \infty] \end{aligned}$$

with parameters as defined earlier and $F_Z(0) = \Pr(N_t = 0) = \frac{(\beta+1)^{\alpha}}{\Gamma(\alpha)}$ for $N = 0$. The exact form of the annual loss cdf is also expressible in closed form,

$$\begin{aligned} F_Z(z) = \sum_{n=1}^{\infty}\frac{(\alpha+n-1)!}{(\alpha-1)!n!}\left(\frac{\beta}{1+\beta}\right)^{\alpha}\left(\frac{1}{1+\beta}\right)^{n}\text{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}}\right) \\ \times\mathbb{I}[\widetilde{\delta}_n < z < \infty] + \frac{(\beta+1)^{\alpha}}{\Gamma(\alpha)}. \end{aligned}$$

One may also consider the binomial and negative binomial models where we assume that the random variable for the total number of events per year are first given by standard binomial and negative binomial processes, before extending these results to doubly stochastic processes. For the binomial processes, we obtain the following result in Theorem 6.13.

Theorem 6.13 (LDA Annual Loss Binomial-Lévy) *The distribution of the annual loss processes Z represented by a compound processes model with LDA structure in which the frequency*

is $N(t) \sim Binomial(M, p)$ *and the severity model* $X_i(t) \sim S(0.5, 1, \gamma, \delta; 0)$*, then the exact density of the annual loss processes can be expressed analytically as a mixture density comprised* α*-stable components with binomial mixing weights for* $N_t > 0$*,*

$$f_Z(z) = \sum_{n=1}^{M} C_n^M (p)^n (1-p)^{M-n} \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z - \widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)} \right) \right]$$
$$\times \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

with $F_Z(0) = \Pr(N_t = 0) = (1-p)^M$ *for* $N = 0$*. The exact form of the annual loss cdf function is also expressible in closed form,*

$$\Pr[Z < z] = F_Z(z) = \sum_{n=1}^{M} C_n^M (p)^n (1-p)^{M-n} \operatorname{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)}} \right)$$
$$\times \mathbb{I}[\widetilde{\delta}_n < z < \infty] + (1-p)^M.$$

For the doubly stochastic binomial-beta process, we can state the following result in Theorem 6.14.

Theorem 6.14 (LDA Annual Loss Doubly-Stochastic-Binomial-Lévy) *The distribution of the annual loss processes* Z *represented by a compound processes model with LDA structure in which the frequency is* $N(t) \sim Binomial(M, p)$*,* $p \sim Beta(\alpha, \beta)$ *and the severity model* $X_i(t) \sim S(0.5, 1, \gamma, \delta; 0)$*, then the exact density of the annual loss processes can be expressed analytically as a mixture density comprised* α*-stable components with beta mixing weights for* $N_t > 0$*,*

$$f_Z(z) = \sum_{n=1}^{M} \int_0^1 C_n^M (p)^n (1-p)^{M-n} \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} dp$$
$$\times \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z - \widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)} \right) \right] \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$
$$= \sum_{n=1}^{M} \frac{M! B(\alpha + n, \beta + M - n)}{(M-n)! n! B(\alpha, \beta)} \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z - \widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)} \right) \right]$$
$$\times \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

with $F_Z(0) = \Pr(N_t = 0) = \frac{B(\alpha,\beta+M)}{B(\alpha,\beta)}$ *for* $N = 0$. *Note that we also denote the standard beta function by* $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$. *The exact form of the annual loss cdf is also expressible in closed form,*

$$\Pr[Z < z] = F_Z(z) = \sum_{n=1}^{M} \frac{M!B(\alpha+n,\beta+M-n)}{(M-n)!n!B(\alpha,\beta)} \text{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}}\right) \times \mathbb{I}[\widetilde{\delta}_n < z < \infty] + \frac{B(\alpha,\beta+M)}{B(\alpha,\beta)}.$$

For the negative binomial process, we obtain the following result in Theorem 6.15.

Theorem 6.15 (LDA Annual Loss Negative Binomial-Lévy) *The distribution of the annual loss processes* Z *represented by a compound processes model with LDA structure in which the frequency is* $N(t) \sim NegativeBinomial(r,p)$ *and the severity model* $X_i(t) \sim S(0.5,1,\gamma,\delta;0)$, *then the exact density of the annual loss processes can be expressed analytically as a mixture density comprised* α*-stable components with negative binomial mixing weights for* $N_t > 0$,

$$f_Z(z) = \sum_{n=1}^{\infty} C_n^{n+r-1}(1-p)^r p^n \sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{(z-\widetilde{\delta}_n)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}\right) \times \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

with $F_Z(0) = \Pr(N_t = 0) = (1-p)^n$ *for* $N = 0$. *The exact form of the annual loss cdf is also expressible in closed form,*

$$\Pr[Z < z] = F_Z(z) = \sum_{n=1}^{\infty} C_n^{n+r-1}(1-p)^r (p)^n \text{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}}\right) \times \mathbb{I}[\widetilde{\delta}_n < z < \infty] + (1-p)^r.$$

For the doubly stochastic negative binomial-beta process, we can state the following result in Theorem 6.16.

Theorem 6.16 (LDA Annual Loss Doubly-Stochastic-Negative-Binomial-Lévy) *The distribution of the annual loss processes* Z *represented by a compound processes model with LDA structure in which the frequency is* $N(t) \sim NegativeBinomial(r,p)$, $p \sim Beta(\alpha,\beta)$ *and the severity model* $X_i(t) \sim S(0.5,1,\gamma,\delta;0)$, *then the exact density of the annual loss processes can be expressed analytically as a mixture density comprised* α*-stable components with beta mixing weights for* $N_t > 0$,

$$f_Z(z) = \sum_{n=1}^{\infty} \int_0^1 C_n^{n+r-1}(1-p)^r p^n \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}dp \times \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z-\widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}\right)\right] \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

$$= \sum_{n=1}^{\infty} \frac{(n+r-1)!B(\alpha+rn,\beta+n)}{(r-1)!n!B(\alpha,\beta)} p^{n(1-r)}(1-p)^{(r-n)}$$
$$\times \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z-\widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)} \right) \right] \mathbb{I}[\widetilde{\delta}_n < z < \infty]$$

with $F_Z(0) = \Pr[N_t = 0] = (1-p)^r$ *for* $N = 0$. *The exact form of the annual loss cdf is also expressible in closed form,*

$$F_Z(z) = \sum_{n=1}^{\infty} \frac{(n+r-1)!B(\alpha+rn,\beta+n)}{(r-1)!n!B(\alpha,\beta)} p^{n(1-r)}(1-p)^{(r-n)} \operatorname{erfc}\left(\sqrt{\frac{\widetilde{\gamma}_n}{2\left(z-\widetilde{\delta}_n\right)}} \right)$$
$$\times \mathbb{I}[\widetilde{\delta}_n < z < \infty] + (1-p)^r.$$

It should be noted that for each of these models as discussed previously, one can obtain a finite sum of mixture components via the practical assumption of an upper bound on the largest number of losses one could encounter for these loss processes. A not unreasonable assumption, given the nature of the loss processes, is considered here, being high consequence losses from heavy-tailed severity models. Alternatively, if this assumption is not suitable, one may truncate the series above as detailed in the following.

Truncation of the infinite series representations can be achieved through a study of the truncation error for the infinite mixture representations of the above-specified single-loss processes. One way to consider where to perform a truncation of these series would be to consider the series expansion of the asymptotic tail expansion for each of the above-specified processes. This will utilize the result present in Lemma 6.11.

Lemma 6.11 (Tail Asymptotics of α-Stable Severity Model) *Given a random variable for the severity of a loss* $X \sim S(\alpha,\beta,\gamma,\delta;0)$, *then as* $x \to \infty$, *one can write the limiting tail distribution*

$$P(X > x) \sim \gamma^{\alpha} c_{\alpha}(1+\beta) x^{-\alpha}, \qquad \text{as } x \to \infty,$$
$$f_X(x|\alpha,\beta,\gamma,\delta;0) \sim \alpha\gamma^{\alpha} c_{\alpha}(1+\beta) x^{-(\alpha+1)}, \qquad \text{as } x \to \infty, \tag{6.218}$$

where $c_{\alpha} = \sin\left(\frac{\pi\alpha}{2}\right)\frac{\Gamma(\alpha)}{\pi}$. *This result follows from Nolan (2015, theorem 1.12).*

One may now use Lemma 6.11 to help to determine a truncation index for the infinite sums as shown in Theorem 6.17. To achieve this, we will first make an approximation in which we consider the index of the summation to be a continuous variable, as demonstrated in, for example, Peters *et al.* (2009b, section 3.1). In addition, one may consider working with the analytic compound processes expression for the tail asymptotic and ensure that after truncation, the error in the finite expansion of the tail asymptotic is bounded to a desired precision. More precisely, the asymptotic tail probability of the compound processes will be established to have a desired approximation error under the truncation derived.

Theorem 6.17 (LDA Annual Loss Finite Series Truncation Point: α-Stable Severity) *Given the LDA structure in which the frequency is distributed generically according to* $N(t) \sim f(\cdot)$

and the severity model $X_i(t) \sim S(0.5, 1, \gamma, \delta; 0)$*, the exact expressions for the tail asymptotic of the resulting infinite mixture is given by*

$$\mathbb{P}\mathrm{r}(Z(t) > z) \sim 2z^{-0.5} c_{0.5} \sum_{n=1}^{\infty} \Pr(N(t) = n) \widetilde{\gamma}_n^{0.5} = C_{\mathit{Tail}} \sum_{n=1}^{\infty} W_n,$$

where $\widetilde{\gamma}_n^{0.5} = \sum_{i=1}^{n} \gamma_i^{0.5} = n|\gamma|^{0.5}$*. These results allows us to determine a unique maximum of* $\frac{d}{dn} \ln(W_n) = 0$ *corresponding to the term* n *with the maximum contribution to the LDA compound processes tail probability for each model. Note that this can be determined trivially for each model by searching for a sign change in the function for each integer value of* n *and so is very efficient to solve because it only requires integer search. The specified equations for each model are given as follows:*

Poisson-*α*-Stable.

Under this model, we have that $W_n = |\gamma|^{0.5} \exp(-\lambda) \frac{\lambda^n}{(n-1)!}$ *which results in the function* $\frac{d}{dn} \ln(W_n) = 0$ *producing* $\ln(\lambda) - \ln(n) - \frac{1}{2n} = 0$*.*

Doubly Stochastic Poisson-Gamma-*α*-Stable.

Under this model, we have that $W_n = |\gamma|^{0.5} \frac{(\alpha+n-1)!}{(\alpha-1)!n!} \left(\frac{\beta}{1+\beta}\right)^{\alpha} \left(\frac{1}{1+\beta}\right)^{n}$ *which results in the function* $\frac{d}{dn} \ln(W_n) = 0$ *producing*

$$\ln(n + \alpha - 1) + \frac{n}{(n + \alpha - 1)} + \frac{(\alpha - 0.5)}{(n + \alpha - 1)} - \ln(n) - 1 - \frac{1}{2n} - \ln(\beta + 1) = 0. \tag{6.219}$$

If we define the maximum value of the terms W_n *obtained as the solutions to the above, by the notation* W_0*, then we can now find* $W_L < W_0 < W_U$ *such that the truncated sum approximation is sufficiently accurate for use in the expression of the annual loss distribution of the compound processes. Hence, we can bound the number of terms in the mixture representations to not change by a precision amount* ε*. This is expressed analytically as*

$$f_Z(z) = \sum_{n=N_L}^{N_U} \Pr(N_t = n) \left[\sqrt{\frac{\widetilde{\gamma}_n}{2\pi}} \frac{1}{\left(z - \widetilde{\delta}_n\right)^{3/2}} \exp\left(-\frac{\widetilde{\gamma}_n}{2\left(z - \widetilde{\delta}_n\right)} \right) \right] \tag{6.220}$$

with N_L *and* N_U *are determined from* $\max(W_L, 1)$ *with* $W_L \leq e^{-37} W_0$ *and* $W_U \leq e^{-37} W_0$*.*

Remark 6.15 *Note that analogous results are easily derived for the other models presented earlier under binomial and negative binomial models.*

6.4.4 CLOSED-FORM LDA MODELS FOR THE TEMPERED α-STABLE SEVERITY FAMILY

In developing the closed-form results for the tempered α-stable models, we first recall some results on infinite divisibility discussed in detail in Chapter 4 which lead to the closure under convolution properties of the TS models. As with the class of α-stable severity models the class of TS severity models are ID and self-decomposable; in other words, they satisfy the condition of closure under convolution presented in Proposition 6.25, see Rosiński (2007, corollary 2.12)

Proposition 6.25 (Closure Under Convolution of Tempered Stable Severity Models (R1)) *If* $\{X_i\}_{i=1}^n$ *are independent ID loss random variables each with a TS distribution* $X_i \sim TS(\alpha, R_i(dx), b_i; R1)$*, then if* $\alpha \in (0,1)$ *and* $\int |x| R(dx) < \infty$*, one has*

$$S_n = \sum_{i=1}^n X_i \sim TS\left(\alpha, \sum_{i=1}^n R_i(dx), \sum_{i=1}^n b_i\right). \tag{6.221}$$

These closure properties for TS models can be used to derive the modified TS and Normal modified TS models in Chapter 4 which were shown to have the property of closure under convolution (infinite divisibility and self-decomposability). If one considers the loss random variable $X \sim TS(\alpha, \beta, \gamma, \delta, \lambda; R6)$ which is built from a strictly positive perfectly skewed stable distribution, then the ID loss random variable X is self-decomposable for the following parameter settings:

1. If $\alpha = \frac{1}{2}$, one obtains the GIG laws which are self-decomposable;
2. If $\nu = -\alpha$, one obtains a class of self-decomposable ID distributions (closed under convolution);
3. If $-(\alpha + \nu) \in \mathbb{J}^+$, one obtains a class of self-decomposable ID distributions (closed under convolution);
4. If $\alpha = \frac{1}{m}$, $\forall m = 2, 3, \ldots$ and $\lambda > 0$ with $\alpha + \nu \in \mathbb{J}^+$, then one obtains a class of self-decomposable ID distributions (closed under convolution).

The following properties of closure under convolution then apply to the following models as was shown in Definition 6.20.

Definition 6.20 (Closure Under Convolution Properties of TS Severity Models) *If* $\{X_i\}_{i=1}^n$ *are independent ID loss random variables each with a TS distribution* $X_i \sim TS(\alpha, R_i(dx), b_i; R1)$*, then if* $\alpha \in (0,1)$ *and* $\int |x| R(dx) < \infty$*, one has*

$$S_n = \sum_{i=1}^n X_i \sim TS\left(\alpha, \sum_{i=1}^n R_i(dx), \sum_{i=1}^n b_i\right). \tag{6.222}$$

The following results apply (see Chapter 5 for definition of different classes of TS distributions):

- $X_i \sim CTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim CTS(\alpha, nc, \lambda_+, \lambda_-, nm)$*;*
- $X_i \sim RDTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim RDTS(\alpha, nc, \lambda_+, \lambda_-, nm)$*;*
- $X_i \sim GTS(\alpha_+, \alpha_-, c_+, c_-, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim CTS(\alpha_+, \alpha_-, nc_+, nc_-, \lambda_+, \lambda_-, nm)$*;*
- $X_i \sim MTS(\alpha, c, \lambda_+, \lambda_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim MTS(\alpha, nc, \lambda_+, \lambda_-, nm)$*;*
- $X_i \sim NMTS(\alpha, c, \lambda, b, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim NMTS(\alpha, nc, \lambda, b, nm)$*;*
- $X_i \sim KRTS(\alpha, k_+, k_-, r_+, r_-, p_+, p_-, m)$ *and the sum of* n *i.i.d. losses becomes* $S_n \sim KRTS(\alpha, nk_+, nk_-, r_+, r_-, p_+, p_-, nm)$.

Given these partial sum results, one can utilize the generic representation of the closed-form LDA model comprised, for instance, Poisson-TS models to obtain the closed-form LDA expressions. Of course, all the variants of the Poisson, binomial, negative binomial, Poisson-Gamma, Poisson-IG mixtures can also be considered for the mixture weights in the distribution. Using these models would require the ability to evaluate and simulate from the TS mixtures, see discussions in Chapter 4.

CHAPTER SEVEN

Single Risk Closed-Form Approximations of Asymptotic Tail Behaviour

In Chapter 6, a set of special sub-families of annual loss loss distribution approach (LDA) models were developed which admitted closed-form exact representations. That is, closed-form representations were obtained for the distribution of the annual loss $Z = X_1 + \cdots + X_N$ represented by a compound process model with LDA structure in which the frequency is $N \sim Poisson(\lambda)$ and the severities are independent and identically distributed (i.i.d.) $X_i(t) \sim F_X(x)$, and N and X_i are independent. The exact distribution of the annual loss processes can be expressed analytically as a mixture distribution comprised convolved distributional components with Poisson mixing weights for $N > 0$,

$$G(z) = \mathbb{P}\mathrm{r}[Z \leq z] = \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[Z \leq z | N = n]\mathbb{P}\mathrm{r}[N = n]$$

$$= \sum_{n=0}^{\infty} p_n F_X^{(n)*}(z). \tag{7.1}$$

Here, $F_X^{(n)*}(z) = \mathbb{P}\mathrm{r}[X_1 + X_2 + \cdots + X_n \leq z]$ is the n-th convolution of the severity distribution $F_X(\cdot)$ calculated recursively as

$$F_{Z_n} = F_X^{(n)*}(z) = \int_0^z F_X^{(n-1)*}(z - x) f_X(x)\, dx$$

with $F_X^{(0)*}(z) = 1$ if $z \geq 0$ and zero otherwise. Note, throughout this chapter we refer to F as the distribution for the severity, unless otherwise indicated by a subscript.

However, in general, outside of these doubly infinitely divisible LDA models, there will exist no exact closed-form solutions for the density and distribution that is easily attainable.

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

Therefore, one can resort to numerical approaches as discussed in Peters *et al.* (2007) and the book length discussions in Shevchenko (2011) and Klugman *et al.* (1998).

In several settings, performing the numerical computation of the annual loss distribution can be computationally demanding, especially if only basic numerical procedures are adopted. This is particularly the case if heavy-tailed models are considered for the severity distribution under integration-based approaches such as in Luo and Shevchenko (2011) and Grübel and Hermesmeier (2000), Panjer-based recursions in Panjer (1981), De Pril (1988), Dickson (1995) or basic Monte-Carlo-based approaches in Shevchenko (2011), McNeil *et al.* (2005a) and Cruz (2002). There have been some more recent developments in reducing the computational costs associated with such evaluations, see discussions in Del Moral *et al.* (2013). However, this still remains a challenge and one would also like a fast comparison for performance.

Alternatively, if the operational risk (OpRisk) model has particular features relating to heavy tailedness, then one can build on the knowledge about such models to utilize the properties of such heavy-tailed models to develop asymptotic representations of the LDA model tails. Often this is the most important component to consider when assessing the impact of a loss process on capital. In the case of random sums such as those of relevance to OpRisk where one considers the tail of the distribution of the compound process,

$$Z = \sum_{i=1}^{N} X_i, \tag{7.2}$$

one needs to consider both the behaviour of the distribution of the severity loss random variables X_i and the distribution of the random number of losses N. The features of such random sums have been studied in numerous places and the following results in Proposition 7.12 characterize the different tail behaviour relationships that can be obtained, see Faÿ *et al.* (2006) and references and discussion in Jessen and Mikosch (2006); for the special case of α-stable random variables in this context, see Feller (2008), Samorodnitsky and Taqqu (1997).

Such asymptotic studies, then naturally lead to the development of closed-form approximations to the quantile function and tail behaviour of the compound processes in the LDA model which is the focus of this chapter. These results then form the building blocks of the results obtained on closed-form approximations to risk measure estimation for value-at-risk (VaR), expected shortfall (ES) and spectral risk measure estimation. The development of such asymptotic tail approximations has been steadily growing in the literature and can be simply classified as follows:

- *'First-Order' Single-Loss Approximations* (*SLA*). The first-order asymptotics were studied in, for example, Teugels and Veraverbeke (1973), Greenwood (1973), von Bahr (1975) and Embrechts and Veraverbeke (1982).
- *'Second-Order' SLAs.* The second-order asymptotics were studied in, for example, Omey and Willekens (1986), Grübel (1987), Mikosch and Nagaev (2001), Willekens (1989) and Baltrūnas (1999).
- *'Higher Order' SLAs.* The higher order expansions were studied in Geluk *et al.* (2000), Borovkov and Borovkov (2002), Barbe and McCormick (2009a) and Barbe *et al.* (2007).

In this chapter, the focus is on an understanding of the tail behaviours that can arise for different LDA models. In particular, we study the tail of the distribution of a sum of a random number of i.i.d. non-negative (loss random variables), that is, the tail of the annual loss model which we will see will depend on the tails of both the number of losses (frequency distribution) and the loss amounts (severity distribution). The following settings will be considered:

1. Conditional on a fixed number of losses an exploration of how the partial sum tail behaviour changes for heavy-tail frequency and light-tail severity is explored.
2. Conditional on a fixed number of losses an exploration of how the partial sum tail behaviour changes for light-tail frequency and heavy-tail severity is explored.
3. In addition, an exploration of how these relationships apply for the compound process random sum settings when there is a weighting associated with the number of terms in the summation is explored.

Such questions have been studied throughout the actuarial literature and have resulted in an ongoing and rich research area, see recent studies in this area that are excellent technical introductions in Borovkov and Borovkov (2010).

To introduce the notion of tail asymptotics and, for instance, the simplest class of these the SLAs one may draw on the works of Barbe and McCormick (2009a) and Albrecher *et al.* (2010) who have studied the behaviour of asymptotic distributions for compound processes comprised sub-exponential distributions which admit regularly varying tail behaviours. To summarize some of the key results presented in this chapter for basic practical applications, we observe the following results which summarize some of the basic asymptotic results that will be presented in this chapter based on different assumptions on the heavy-tailed features of the frequency and severity models.

In this chapter, we will study in more depth the simple results later and explain their origin and their extensions for practitioners including the higher order approximations of such tail behaviour of the annual loss distribution. The following results for compound process tail behaviour can be shown to hold for first-order approximations under different assumptions relating to the frequency and severity models utilized in the LDA structure:

1. ***Heavy-Tailed Frequency, Light-Tailed Severity.*** Assume that the counting random variable N is regularly varying with index $\beta \geq 0$. If $\beta = 1$, then assume that $\mathbb{E}[N] < \infty$. Let the i.i.d. loss random variables (X_i) have finite mean $\mathbb{E}[X_i] < \infty$ and tail probability with asymptotic condition $\Pr[X_i > x] = o(\Pr[N > x])$, then asymptotically one has the equivalence for the compound processes given as follows:

$$\Pr[Z_N > x] \sim (\mathbb{E}[X_1])^{\beta} \Pr[N > x]. \tag{7.3}$$

2. ***Heavy-Tailed Frequency, Heavy-Tailed Severity.*** Assume that $\Pr[N > x] \sim c\Pr[X_1 > x]$ for some $c > 0$ and X_1 is regularly varying with an index $\rho \geq 1$ and with finite mean $\mathbb{E}[X_1] < \infty$. Then one has the asymptotic equivalence

$$\Pr[Z_N > x] \sim \Pr[X_1 > x] (\mathbb{E}[N] + c(\mathbb{E}[X_1])^{\rho}). \tag{7.4}$$

3. ***Light-Tailed Frequency, Heavy-Tailed Severity (Case 1).*** Assuming that the distribution of X_i is regularly varying with positive index $\rho > 0$ and the mean number of losses is finite $\mathbb{E}[N] < \infty$ and further more that the following asymptotic condition is satisfied $\Pr[N > x] = o(\Pr[X_1 > x])$. Then the following asymptotic equivalence holds

$$\Pr[Z_N > x] \sim \mathbb{E}[N] \Pr[X_1 > x]. \tag{7.5}$$

4. ***Light-Tailed Frequency, Heavy-Tailed Severity (Case 2).*** If the compound process Z_N is regularly varying with index $\rho > 0$ and the moment bound on the severity distribution $\mathbb{E}[X_1] < \infty$ applies, furthermore, one has the asymptotic condition $\Pr[X_1 > x] = o(\Pr[Z_N > x])$. In the case in which $\rho = 1$ and $\mathbb{E}[Z_N] = \infty$, assume that asymptotically

$x\mathbb{P}\mathrm{r}[X_1 > x] = o(\mathbb{P}\mathrm{r}[Z_N > x])$. Then N is regularly varying with index ρ and one has the following asymptotic equivalence for the annual loss:

$$\mathbb{P}\mathrm{r}[Z_N > x] \sim \mathbb{E}[Z_N > x] \sim (\mathbb{E}[X_1])^{\rho}\mathbb{P}\mathrm{r}[N > x]. \tag{7.6}$$

5. ***Light-Tailed Frequency, Heavy-Tailed Severity (Case 3).*** If the compound process Z_N is regularly varying with index $\rho > 0$ and the moment bound on the frequency distribution, for some constant $\delta > 0$, given by $\mathbb{E}[N^{1\vee(\rho+\delta)}] < \infty$, then X_1 is regularly varying with index ρ and the following asymptotic equivalence holds

$$\mathbb{P}\mathrm{r}[Z_N > x] \sim \mathbb{E}[N]\mathbb{P}\mathrm{r}[X_1 > x]. \tag{7.7}$$

Note, the case of light tailed frequency and light tailed severity is also considered and will be discussed later in this chapter. Such cases are typically not of interest to the models in OpRisk, but we present some standard asymptotics for this case for completeness and comparison purposes. The result derived involves a standard saddle point approximation for the distribution of a partial sum and a random sum. The remainder of this chapter will be devoted to an understanding of these results and their extensions to higher order loss approximations. As such, this chapter is structured as follows:

1. Study of asymptotics for partial and compound processes:
   (a) Light-tailed frequency and heavy-tailed severity models;
   (b) Heavy-tailed frequency and light-tailed severity models;
   (c) Heavy-tailed frequency and heavy-tailed severity models;
2. First-order, second-order and higher order asymptotic tail expansions for the annual loss distribution (LDA) models.

Throughout the following sections, it will be assumed that one has a basic knowledge of the definitions and properties of the different classes of heavy-tailed distributional classifications that were discussed in Chapter 4, Chapter 5 and Chapter 6. Next we consider how to utilize these concepts of characterization of heavy-tailed severity models in OpRisk LDA structures to understand the partial sum and compound process asymptotic tail behaviours of the annual loss.

7.1 Tail Asymptotics for Partial Sums and Heavy-Tailed Severity Models

Throughout this section, we consider the question of what can be stated about the asymptotics of the tail of the sum of n losses (partial sum) when the tail properties of the distribution from which the losses were drawn is assumed known. This corresponds to the case in which one models the annual loss conditional upon $N = n$ losses occurring and studies the tail asymptotics and, therefore, types of capital requirements that may be encountered if this many losses occurred and the severity model had particular tail attributes. In this context, there will be four basic settings of interest to OpRisk modeling:

1. the case when the number of terms in the partial sum n is finite and the n loss random variables that are aggregated are independent;

2. the case when the number of terms in the partial sum n is finite and the n loss random variables that are aggregated are dependent in some manner;
3. the case when the number of terms in the partial sum n is finite and the n loss random variables that are aggregated are independent and each is weighted in some manner (such as would occur when insurance is incorporated or perhaps key-risk indicators are considered;
4. the case when the number of terms in the partial sum n is infinite (or approaching infinity) and the n loss random variables that are aggregated are independent (perhaps also weighted).

To address these different settings, this section is structured in two parts, the first looks at the asymptotic results that can be obtained for the partial sum of n losses when n is finite and the second case considers the special case when n is infinite, with the aim to progressing towards an understanding of results for compound (random) sum settings.

7.1.1 PARTIAL SUM TAIL ASYMPTOTICS WITH HEAVY-TAILED SEVERITY MODELS: FINITE NUMBER OF ANNUAL LOSSES $N=n$

In this subsection, we will primarily be concerned with partial sums (conditional upon a known number of losses $N = n$) having occurred and we will aim to further characterize some important results regarding the tail asymptotics for such partial sums $Z_n = X_1 + \cdots + X_n$ which are comprised i.i.d. random losses X_i. The first of these results to consider is given in Proposition 7.1, see Geluk (1992).

Proposition 7.1 (General Partial Sum Distributional Lower Bounds) *Consider a severity distribution $\overline{F}(x) > 0$ for losses that are strictly positive $x > 0$, then the following two results on lower bounds can be shown to hold for any $n \geq 1$:*

1. *The greatest lower bound on the asymptotic behaviour of the ratio of the partial sum tail to the severity model tail is given by the number of losses accumulated according to*

$$\liminf_{x\to\infty} \frac{\overline{F_X^{(n)*}}(x)\overline{F}_X}{\overline{F}_X(x)} \geq n. \tag{7.8}$$

2. *This lower bound can be further refined using powers of the severity model, where for some $k \in [1, n-1]$, one can show the following greatest lower bound on the asymptotic tail behaviour*

$$\liminf_{x\to\infty} \frac{\overline{F_X^{(n)*}}(x) + \sum_{j=1}^{k} \frac{n!}{(n-j)!j!}(-1)^j \overline{F}_X(x)^j}{\overline{F}_X(x)^{k+1}} \geq (-1)^k \frac{n!}{(n-k-1)!(k+1)!}. \tag{7.9}$$

Remark 7.1 *These results are useful as typically one has good knowledge of the severity distribution tails but not perhaps the partial sums asymptotic tail behaviour, as such these bounds allow one to understand how the partial sum tail behaviour is affected, lower bounded by the behaviour of the actual severity model or powers of the severity tail as is the case for the second result.*

One can make the additional assumption that the losses are i.i.d. with a severity distribution which is sub-exponential $\overline{F}_X \in \mathscr{F}$, such that

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_1 + X_2 > x]}{\mathbb{P}\mathrm{r}[X_1 > x]} = 2, \tag{7.10}$$

with tail probabilities $\mathbb{P}\mathrm{r}[X_i > x] = 1 - F_X(x) = \overline{F}_X(x)$. Hence, a distribution F_X on the positive half-line is sub-exponential $F_X \in \mathscr{F}$ if, as $x \to \infty$, one has the asymptotic relationship

$$\overline{F}_{Z_2}(X) = \overline{F_X^{(2)*}}(x) \sim 2\overline{F}_X(x). \tag{7.11}$$

Then it can, therefore, be shown that for any $n \geq 1$, one has asymptotically, as $x \to \infty$,

$$\overline{F_X^{(n)*}}(x) \sim n\overline{F}_X(x) \quad \text{as } x \to \infty. \tag{7.12}$$

This means that given a sum of n independent random variables $X_1, \ldots, X_n$ with common distribution F_X, one has the following probabilistic interpretation for sub-exponential distributed random variables (often referred to as the 'big jump'),

$$\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x] \sim \mathbb{P}\mathrm{r}[\max(X_1, \ldots, X_n) > x], \text{ as } x \to \infty. \tag{7.13}$$

This result arises from the condition for sub-exponential i.i.d. random variables which is known as the convolution root closure, defined in Lemma 7.1 below.

Lemma 7.1 (Convolution Root Closure of Sub-exponential Distributions) *Assume that the partial sum $Z_n = \sum_{i=1}^n X_i$ is regularly varying with index $\rho \geq 0$ with each X_i being i.i.d. with positive support. Then for all $i \in \{1, \ldots, n\}$, the X_i's are regularly varying also with index ρ and the following asymptotic equivalence as $x \to \infty$ holds*

$$\mathbb{P}\mathrm{r}[Z_n > x] \sim n\mathbb{P}\mathrm{r}[X_1 > x], \ \forall n \geq 1. \tag{7.14}$$

Remark 7.2 *The intersection between the family of dominantly varying, subversively varying and sub-exponential-tailed distributions is contained in the family comprised of the intersection between the dominantly varying tailed functions and the long-tailed functions. Furthermore, these sub-families formed from the intersections are themselves contained in the family of sub-exponential models which is itself contained in the family of long-tailed distributions. A detailed discussion on these families is provided in Chapter 1 and Chapter 2.*

From the properties of distributions which are tail balanced and regularly varying distributions, one can also show the result in Lemma 7.2, see Embrechts *et al.* (1997, appendix 3.26).

Lemma 7.2 *Consider an i.i.d. sequence of loss random variables X_i for $i \in \{1, 2, \ldots, n\}$ with distribution function F which is regularly varying with index $\rho \geq 0$ and satisfies the tail balance condition for some $p + q = 1$ in Definition 3.3. Then for any real constants $c_i \in \mathbb{R}$ and integer $n \geq 1$, one can show*

$$\mathbb{P}\mathrm{r}[c_1X_1 + \cdots + c_nX_n > x] \sim \mathbb{P}\mathrm{r}[|X_1| > x] \sum_{i=1}^n [p(c_i^+)^\rho + q(c_i^+)^\rho], \tag{7.15}$$

where $c_i^+ = 0 \vee c_i$ and $c_i^- = 0 \vee (-c_i)$.

Note, in general, except when $n = 2$, it is not true that given a linear combination $c_1X_1 + c_2X_2 + \cdots + c_nX_n$ which is regularly varying with index ρ for an i.i.d. sequence of losses (X_i), then this does not imply that X_1 is from a distribution which is regularly varying.

It is now beneficial to recall the definitions of O-regular variation, that was presented in Chapter 3, which is recalled below for convenience. First, defining for $\lambda > 0$ the notations

$$f^*(\lambda) = \limsup_{x\to\infty} \frac{f(\lambda x)}{f(x)} \tag{7.16}$$

and

$$f_*(\lambda) = \liminf_{x\to\infty} \frac{f(\lambda x)}{f(x)}, \tag{7.17}$$

with the relationship that $f_*(\lambda) = \frac{1}{f^*(1/\lambda)}$. One can then recall the definition of the O-regular variation given below in Definition 7.1.

Definition 7.1 (Extended and O-Type Regular Variation) *The class of extended regularly varying functions is the set of positive measurable functions $f \in ER$ satisfying for some constants c, d the relationship*

$$\lambda^d \le f_*(\lambda) \le f^*(\lambda) \le \lambda^c, \ \forall \lambda \ge 1. \tag{7.18}$$

The class of O-regularly varying functions is the set of positive measure functions $f \in OR$ satisfying

$$0 < f_*(\lambda) \le f^*(\lambda) < \infty, \ \forall \lambda \ge 1. \tag{7.19}$$

Another way to define a distribution function as O-regularly varying is given in Definition 7.2, see Geluk (1992).

Definition 7.2 (O-Regular Variation of Distribution Functions) *A distribution function F that is concentrated on $[0, \infty)$ is of O-regular variation $F \in OR(\alpha)$ if there exists a positive function α which is O-regularly varying, that is, measurable, eventually positive and satisfying for $x > 0$ the condition*

$$\lim_{t\to\infty} \frac{\alpha(tx)}{\alpha(t)} < \infty, \tag{7.20}$$

where $\alpha(x) = o(\overline{F}_X(x))$ as $x \to \infty$ and

$$\limsup_{x\to\infty} \frac{\overline{F}_X\left(\frac{1}{2}x\right) - \overline{F}_X(x)}{\alpha(x)} < \infty. \tag{7.21}$$

Using these definitions, one can show the following bounds for the partial sum of n-losses, each of which is assumed to be i.i.d. from a severity distribution $F_X \in OR(\alpha)$, given in Proposition 3.5, see Geluk (1992).

Proposition 7.2 (Partial Sum Bounds for O-Regularly Varying Severities) *If a distribution F_X satisfies that it is a member of the O-regularly varying class $F_X \in OR(\alpha)$, then there exists constants δ_n for $n \ge 2$ such that the following bound on the tail of the distribution of the partial sum in terms of the tail of the severity distribution is obtained*

$$\overline{F_X^{(n)*}}(x) - n\overline{F}_X(x) \le \delta_n \alpha(x), \ \ x > 0. \tag{7.22}$$

Using this result, in Geluk (1992) it was shown that the following asymptotic big-O order of approximation for the n-fold convolution approximation of the partial sum distribution tail by the n-th power of the severity distribution tail is given as presented in Proposition 7.3.

Proposition 7.3 (Partial Sum Tail Distribution Approximation via Severity Tail Distribution Powers) *Consider the distribution function for the loss random variables $F_X(x)$ that satisfies the condition*

$$\limsup_{x\to\infty} \frac{\overline{F}_X\left(\frac{1}{2}x\right) - \overline{F}_X(x)}{\overline{F}_X(x)^3} < \infty, \tag{7.23}$$

then one has the order of approximation

$$\overline{F_X^{(n)*}}(x) - F_X^n(x) = O(\overline{F}_X(x)^3). \tag{7.24}$$

Omey and Willekens (1986) also note in their study of the remainder terms in a first-order LDA asymptotic error analysis that the following result for partial sums may be obtained, as given in Proposition 7.1 for regularly varying severity distribution functions.

Theorem 7.1 (Remainder Analysis for Partial Sum Approximations: RV Case) *If one considers an OpRisk LDA loss model with a severity distribution that satisfies that the density exists and is in the class of regularly varying functions $f_X \in RV_{-\alpha}$ for some $\alpha \in [1,2]$, then the remainder term of the partial sum of n-losses given by $R(x) = \overline{F}_{Z_n}(x) - n\overline{F}_X(x)$ will satisfy the following asymptotics*

$$\lim_{x\to\infty} \frac{\overline{F}_{Z_n}(x) - n\overline{F}_X(x)}{f_X(x)\int_0^x \overline{F}_X(y)\,dy} = \begin{cases} c(\alpha)\dfrac{n!}{2(n-2)!}, & \text{if } \alpha \in (1,2), \\ \dfrac{n!}{(n-2)!}, & \text{if } \alpha = 2, \end{cases}$$

$$\lim_{x\to\infty} \frac{\overline{F}_{Z_n}(x) - n\overline{F}_X(x)}{\overline{F}_X(x)^2} = -\frac{n!}{2(n-2)!} \quad \text{if } \alpha = 1, \tag{7.25}$$

with function $c(\alpha)$ given by

$$c(\alpha) = \frac{(2-\alpha)\Gamma(2-\alpha)^2}{(\alpha-1)\Gamma(3-2\alpha)}. \tag{7.26}$$

Geluk (1992) also provides an analogue of the famous convolution root closure condition for sub-exponential distributions for the case of O-regularly varying distributions as given in Proposition 7.4. This shows how the O-regular variation properties present in severity model's tail features are preserved under n-fold convolution for the distribution of the partial sum of n losses.

Proposition 7.4 (Convolution Root Closure of O-Regular Variation Distributions) *Consider a severity distribution F_X that satisfies that its n-fold convolution with itself ${F_X}^{(n)*}$ is of O-regular variation $F_X^{(n)*} \in OR(\alpha)$ for some $n \geq 2$ and the tail of the n-fold convolution satisfies $(F_X^{(n)*}(x))^2 = O(\alpha(x))$ as $x \to \infty$, then one has that the severity distribution is also of O-regular variation $F_X \in OR(\alpha)$. Hence, the severity distribution F_X is closed under convolution.*

If one considers a finite sum of independently distributed random variables $X_i \sim F_i$, denoted by $Z_n = \sum_{i=1}^n X_i$ where the distributions F_i are regularly varying with possible different indexes, then a result characterizing each of their contributions to the right tail of the

distribution of $\overline{F}_{Z_n}(x)$ is given in Lemma 7.3, see Jessen and Mikosch (2006, lemma 3.1). This result will prove to be important for studying higher order tail asymptotics (decompositions 2 and 3) proposed by Albrecher *et al.* (2010).

Lemma 7.3 *Consider a finite sum of independently distributed random variables $X_i \sim F_i$, denoted by $Z_n = \sum_{i=1}^n X_i$. Furthermore, assume that the distributions of random variables $X_i \sim F_i$ are regularly varying, possibly with different indexes and that the random variable $|X_1|$ has a distribution (denoted by F) which is regularly varying with index $\rho \geq 0$. Assume that the random variables $X_1, \ldots, X_n$ each have distributions satisfying the condition that*

$$\begin{aligned} \lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_i > x]}{\overline{F}(x)} &= c_i^+, \\ \lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_i \leq -x]}{\overline{F}(x)} &= c_i^-, \end{aligned} \tag{7.27}$$

for non-negative constants c_i^+ and c_i^- and the pairwise conditions that

$$\begin{aligned} \lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_i > x, X_j > x]}{\overline{F}(x)} &= \lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_i \leq -x, X_j > x]}{\overline{F}(x)} \\ &= \lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_i \leq -x, X_j \leq -x]}{\overline{F}(x)} = 0, \ \forall i \neq j. \end{aligned} \tag{7.28}$$

If these conditions are satisfied, the following holds for the tail distribution of the partial sum $\overline{F}_{Z_n}$,

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[Z_n > x]}{\overline{F}(x)} = \sum_{i=1}^n c_i^+ \tag{7.29}$$

and

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[Z_n \leq -x]}{\overline{F}(x)} = \sum_{i=1}^n c_i^-. \tag{7.30}$$

The implications of this are that for independent loss random variables (positive support) which have a severity distribution which is regularly varying, the partial sum has the following asymptotic order of approximation as $x \to \infty$,

$$\overline{F}_{Z_n}(x) \sim \sum_{i=1}^n \mathbb{P}\mathrm{r}[X_i > x]. \tag{7.31}$$

In Ng *et al.* (2002), it was shown that for long-tailed severity distributions, one could also obtain partial sum asymptotics. First, we recall the definition of a severity distribution which is long tailed, see Chapter 3.

Definition 7.3 (Long-Tailed Distribution) *A severity distribution function F_X is said to be long tailed if it satisfies the asymptotic condition that*

$$\lim_{x\to\infty} \frac{\overline{F}_X(x-y)}{\overline{F}_X(x)} = 1, \tag{7.32}$$

for all constants $y \in \mathbb{R}$. Equivalently, one can state this by saying that $\overline{F}_X(x-y)$ is asymptotically equivalent to $\overline{F}_X(x)$, that is, $\overline{F}_X(x-y) \sim \overline{F}_X(x)$.

If one has a partial sum of n-losses which are i.i.d. with a severity distribution which is long tailed, then one can derive the following results in Proposition 7.5, see Ng *et al.* (2002, theorem 2.1).

Proposition 7.5 (Partial Sum Tail Asymptotics for Long-Tailed Severity Models) *Consider the partial sum of n i.i.d. losses with severity distribution $X_i \sim f_X$ which is long tailed. Then one has, for each $n \geq 1$, the following asymptotic result*

$$F_{Z_n}(x) \sim \mathbb{P}\mathrm{r}\left[\max_{1\leq k\leq n} Z_k > x\right]. \tag{7.33}$$

This result generalizes the sub-exponential result discussed previously. Next we consider the setting in which the number of terms in the partial sum may be infinite and address the question relating to what are the conditions required for the convergence of such infinite sequences of loss random variables.

7.1.2 PARTIAL SUM TAIL ASYMPTOTICS WITH HEAVY-TAILED SEVERITY MODELS: LARGE NUMBERS OF LOSS EVENTS

Typically, in OpRisk, one will consider either finite weighted series of i.i.d. random variables, $\sum_{i=1}^{n} c_i X_i$ as above, or the case of infinitely weighted series where $n = \infty$. Practically, one may bound the total number of losses in an LDA model to an upper bound amount for a year, although the standard practice in OpRisk is to utilize a Poisson distribution for the number of losses which clearly allows a non-zero probability of very large numbers of losses $n \to \infty$ with probability decaying at a rate of $\lambda^n/n!$ as $n \to \infty$. So, even in practical settings, it is important to consider large and infinite $N = n$ partial summations of loss random variables when assessing the properties of the tail of the annual loss distribution. In addition, as was discussed in the closed-form LDA models in Chapter 6, there are some loss processes in OpRisk with very large numbers of losses that occur throughout the year, such as credit card frauds. In such cases, the following results will also be of relevance.

7.1.2.1 Preliminary Background Results for Partial Sums. In the case of an infinitely weighted sum of i.i.d. heavy-tailed loss random variables, for instance, with severity which is regularly varying, the question of regular variation of the distribution of the sum $\sum_{i=1}^{\infty} c_i X_i \sim G(x)$ is only sensibly addressed if the series is convergent almost surely (a.s.). To understand the convergence of such a series, it was studied in Durrett (2010) as shown in Theorem 7.2.

Theorem 7.2 (Kolmogorov's Three Series Theorem) *Suppose $X_1, X_2, \ldots$ are independent, and define $Y_n = X_n \mathbb{I}[|X_n| \leq 1]$. Then the sum $\sum_{n=1}^{\infty} X_n < \infty$ a.s. if and only if*

1. $\sum_{n=1}^{\infty} \mathbb{P}\mathrm{r}[|X_n| > 1] < \infty$;
2. $\sum_{n=1}^{\infty} \mathbb{E}[Y_n] < \infty$; *and*
3. $\sum_{n=1}^{\infty} \mathbb{V}\mathrm{ar}(Y_n) < \infty$.

Other basic results of importance for understanding the behaviour of the sequences of partial sums, as the sequence goes to infinity, will be of direct significance later, especially when studying the impact of dependence on asymptotics in partial sums and compound process. These results involve the understanding of the distributions of the finite sequences of partial

sums, that is, a characterization or representation of their characteristic functions. In general cases in which the random variables being summed can be positive or negatively valued, this type of analysis is known as fluctuation analysis. A key feature of such results is to consider analysis of the exact distribution of variables defined by a finite number of the losses in the partial sums Z_n; this is contrast to limit theorem type results which study asymptotics based on domains of attraction and consideration of convergence to a particular distributional family as n grows large.

Key results on the existence and representations of the characteristic function for such infinite sequences of weighted sums was also considered in Pollaczek (1957), Pollaczek (1975), De Smit (1973) and Trinidad (1990). An elegant summary of some of these results was provided in Baxter (1961). The following results relate to a rich literature that has been developing for several decades on distributions of sequences of partial sums and maxima of such sequences; as noted, this area of probability is sometimes known as fluctuation analysis and also relates to distributions of records in statistics and extreme value theory (EVT).

We begin with some basic notations to demonstrate the significance of such results in this section. Given an infinite sequence of i.i.d. loss random variables $\{X_1, X_2, \dots\}$ with a common severity distribution $f_X(x)$, then the study of the random variables given by, for instance, the k-th order statistics for the maximum of all the partial sums in the sequence, given for all $n \geq 1$ by

$$R_{(k,n)} := \overset{(k)}{\max_{1 \leq i \leq n}} \{Z_i\}, \tag{7.34}$$

where $Z_i = \sum_{j=1}^{i} X_j$ and the min and max order statistics would, therefore, be given for $k = 1$ and $k = n$, respectively, by

$$\begin{aligned} R_{(1,n)} &= \overset{(1)}{\max_{1 \leq i \leq n}} \{Z_i\} = \max_{1 \leq i \leq n} \{Z_i\}, \\ R_{(n,n)} &= \overset{(n)}{\max_{1 \leq i \leq n}} \{Z_i\} = \min_{1 \leq i \leq n} \{Z_i\}. \end{aligned} \tag{7.35}$$

The study of the distribution and tail behaviour of the random variables $R_{(k,n)}$ was performed by Pollaczek (1952), Pollaczek (1975) and Wendel (1960). Each author discovered through different mechanisms a generalization of Spitzer's identity now known as the Pollaczek-Spitzer-Wendell (PSW) identity, see Definition 7.4. This identity provides an understanding of the distribution of order statistics for the maximum of random sums in an infinite sequence and relates to the properties of the well-known Wiener–Hopf factorization, see discussions in De Smit (1973). As pointed out in De Smit (1973, theorem 1), the distribution of $R_{(k,n)}$ random variables is identical to that of the distribution given by random variables (for i.i.d. sequences) according to the following definition:

$$Z_{(k,n)} = \overset{(k)}{\max_{1 \leq i \leq n}} (Z_n - Z_{n-1}), \quad n = 1, 2, \dots; \quad k = 2, 3, \dots, n. \tag{7.36}$$

Furthermore, it was proven in this theorem that for all $n \geq 0$,

$$Z_{(1,n+1)} = Z^{+}_{(1,n)} + X_{n+1}, \tag{7.37}$$

where $x^+ := \max(0, x)$ and $x^- = \min(0, x)$ and that

$$Z_{(k,n+1)} = Z^{+}_{(k,n)} + Z^{-}_{(k,n)} + X_{n+1}, \tag{7.38}$$

for all $k \in \{1, 2, \dots, n+1\}$. Note that this is not so significant in OpRisk settings where loss random variables are always strictly non-negative-valued random variables.

One additional definition will also be relevant, where the function $\varphi(q, r, \rho_1, \rho_2)$ is given, for $|q| < 1$, $|r| < 1$, $\mathscr{R}\mathrm{e}\{\rho_1\} \geq 0$ and $\mathscr{R}\mathrm{e}\{\rho_2\} \geq 0$, by

$$\varphi(q, r, \rho_1, \rho_2) := \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} q^n r^{k-1} \mathbb{E}[\exp(-\rho_1 Z^+_{(k,n)} - \rho_2 Z_n)]. \tag{7.39}$$

Note that for $\mathscr{R}\mathrm{e}\{\rho\} = 0$, one has $\Gamma(\rho) = \int_{-\infty}^{\infty} \exp(-\rho x) d f_X(x)$.

With these definitions and the results of (De Smit 1973, theorem 1), one may to now present the identity of PSW. This identity is instrumental in understanding the properties of the distribution for records and order statistics for partial sums in a sequence of an infinite numbers of losses as constructed above. Note that the results are simplified in cases in which the random variables are strictly positive.

Definition 7.4 (Pollaczek–Spitzer–Wendell Identity) *The PSW identity for the function* $\varphi(q, r, \rho_1, \rho_2)$ *is given by*

$$\begin{aligned}
\varphi(q, r, \rho_1, \rho_2) &:= \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} q^n r^{k-1} \mathbb{E}\left[\exp\left(-\rho_1 Z^+_{(k,n)} - \rho_2 Z_n\right)\right] \\
&= \frac{1}{(1-r)[1 - q\Gamma(\rho_2)]} \\
&\quad \times \exp\left\{ \sum_{n=1}^{\infty} \frac{q^n}{n} (r^n - 1) \mathbb{E}\left[\exp(-\rho_2 Z_n)\left(1 - \exp(-\rho_1 Z_n^+)\right)\right] \right\},
\end{aligned}$$

for $|q| < 1$, $|r| < 1$, $\mathscr{R}\mathrm{e}\{\rho_1\} \geq 0$ *and* $\mathscr{R}\mathrm{e}\{\rho_2\} = 0$.

Remark 7.3 *To understand where this identity applies, it is important to recognize that it will be associated to the characteristic function of the order statistics for sequences of random variables such as those given by the following partial sums:*

$$\begin{aligned}
Z_{(k,n)} &= \overset{(k)}{\max} \left\{ X_1, X_1 + X_2, \ldots, \sum_{i=1}^{n} X_i \right\}, \\
Z^+_{(k,n)} &= \max\{Z_{(k,n)}, 0\}, \\
Z^-_{(k,n)} &= \min\{Z_{(k,n)}, 0\}, \quad k = 1, 2, 3, \ldots, n.
\end{aligned}$$

See further discussions and extensions of the above identity to characteristic functions of such quantities in Takács (1977).

These are just some examples of the fluctuation analysis results that are available, where a key feature of such results is to consider analysis of the exact distribution of variables defined by a finite number of losses in the partial sums Z_n. This is in contrast to limit theorem type results based on domains of attraction for the convergence to a particular distribution for large n. The above general PSW identity is a general result that relates to the identity obtained by Spitzer and independently by Pollaczek. This identity showed that the distribution of the first-order statistic, given by the max for the sequences of partial sums, which may not be ordered in general for random variables with positive and negative supports given according to

$$R_{(1,n)} = \overset{(1)}{\max_{1 \leq i \leq n}} \{Z_i\} = \max\{Z_1, Z_2, \ldots, Z_n\}$$
$$= \max\left\{X_1, X_1 + X_2, \ldots, \sum_{i=1}^{n} X_i\right\}, \tag{7.40}$$

was given by the identity of the form of a weighted summation of quantities resembling the characteristic function of the maxima of the partial sums, with the exception that the expectation is taken with regard to the severity model distributions for each partial sum $F_X^{(k)*}(x)$ according to

$$\sum_{n=0}^{\infty} \lambda^n \int \exp\left(\mathrm{i}\theta R_{(1,n)}\right) dF_X^{(n)*}(x)$$
$$= \exp\left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k} \left(\int_{\{Z_k > 0\}} \exp\left(\mathrm{i}\theta Z_k\right) dF_X^{(k)*}(x) + \mathbb{P}[Z_k \leq 0]\right)\right], \tag{7.41}$$

see discussions in Baxter (1961). Note that in the case of OpRisk, we would only care about the case in which the $X_i \geq 0$, where in such cases we consider the sequences of increasing partial sums given by the order statistics for the maxima of the partial sums $R_{(1,n)} \geq R_{(2,n)} \geq \cdots \geq R_{(n,n)}$. This then leads to Wendell's contribution which involves the identity linking the characteristic functions of the records $R_{(m,n)}$ to an expression in terms of the characteristic function of the original partial sums, given by

$$\sum_{n=0}^{\infty} \lambda^n \sum_{m=0}^{n} U^m \int \exp\left(\mathrm{i}\theta R_{(m,n)}\right) dF_{R_{(m,n)}}$$
$$= \exp\left[\sum_{k=1}^{\infty} \frac{\lambda^k}{k} \left(\int_{\{Z_k > 0\}} \exp\left(\mathrm{i}\theta Z_k\right) dF_X^{(k)*}(x) + U^k \mathbb{P}[Z_k > 0]\right.\right.$$
$$\left.\left. + U^k \int_{\{Z_k \leq 0\}} \exp\left(\mathrm{i}\theta Z_k\right) dF_X^{(k)*}(x) + \mathbb{P}[Z_k \leq 0]\right)\right]. \tag{7.42}$$

A general framework for these results and others was developed by Baxter, see Baxter (1961) for details.

As previously mentioned, the only result of interest to OpRisk in these general settings is the restriction to the case in which $\mathbb{P}\mathrm{r}[X_i \geq 0] = 1$, and therefore, one has the simplifications of the above results to $\mathbb{P}\mathrm{r}[Z_n > 0] = 1$ for all $n \geq 1$. In addition, the expectations in each of these identities is taken with respect to the distributions of the loss random variables creating the partial sum; if they are i.i.d. as mentioned earlier, this would of course be given by the standard n-fold convolution as stated in the examples mentioned earlier. However, this could be more general if one had dependence between the loss random variables.

7.1.2.2 Asymptotics for Heavy-Tailed Severity Models in Partial Sums. If the index of regular variation of each of the i.i.d. random variables X_i is $\rho > 2$, then each of these variables have finite variance and the required necessary and sufficient conditions for the a.s. convergence of the infinite series become $\mathbb{E}[X_i] = 0$ and $\sum_{i=1}^{\infty} c_i^2 < \infty$. When one considers this theorem extended to the infinitely weighted series and considers the requirements on the sequence (c_i) and (X_i), the conditions in Lemma 7.4 show the requirements for convergence, see details in Mikosch and Samorodnitsky (2000).

Lemma 7.4 (Tail Asymptotics for Partial Sums with Large Numbers of Claims) *Consider an infinite i.i.d. sequence of random losses* (X_i) *which have a regularly varying distribution with index* $\rho > 0$ *which satisfies the tail balance condition, for some* $(p + q = 1)$*, in Definition 3.3 and an infinite sequence* (c_i) *of real weights. If any one of the following conditions is satisfied by these sequences:*

1. *index of regular variation* $\rho > 2$, $\mathbb{E}[X_1] = 0$ *and* $\sum_{i=1}^{\infty} c_i^2 < \infty$*;*
2. *index of regular variation* $\rho \in (1, 2]$, $\mathbb{E}[X_1] = 0$ *and* $\sum_{i=1}^{\infty} |c_i|^{\rho-\varepsilon} < \infty$ *for some* $\varepsilon > 0$*; or*
3. *index of regular variation* $\rho \in (0, 1]$ *and* $\sum_{i=1}^{\infty} |c_i|^{\rho-\varepsilon} < \infty$ *for some* $\varepsilon > 0$*;*

then the tail probability for $Z = \sum_{i=1}^{\infty} c_i X_i$ *satisfies asymptotically*

$$\mathbb{P}\mathrm{r}[Z > x] \sim \mathbb{P}\mathrm{r}[|X_1| > x] \sum_{i=0}^{\infty} [p(c_i^+)^\rho + q(c_i^-)^\rho]. \tag{7.43}$$

In general, the context of OpRisk will always consider purely positively weighted sequences of i.i.d. loss random variables, in which case the convergence of the partial sums from an infinite sequence of positive weighted sums holds according to Theorems 7.3 and 7.4, see Sundt and Vernic (2009, theorems 1.9 and 1.10).

Theorem 7.3 *If* s *is a non-negative integer and* $\{x_n\}_{n=s}^{\infty}$ *is a sequence of positive numbers such that the* $\lim_{n\to\infty} x_n/x_{n-1} = c$ *exists, then the infinite series* $\sum_{n=s}^{\infty} x_n$ *is convergent if* $c < 1$ *and divergent if* $c > 1$*.*

Alternatively, one can consider the specification for a convergence of the partial sums given in Theorem 7.4 which provides alternate conditions on the convergence of the partial sums for values of c.

Theorem 7.4 *If* s *is a non-negative integer and* $\{x_n\}_{n=s}^{\infty}$ *is a sequence of positive numbers such that the* $\lim_{n\to\infty} n\left(1 - x_n/x_{n-1}\right) = c$ *exists, then the infinite series* $\sum_{n=s}^{\infty} x_n$ *is convergent if* $c > 1$ *and divergent if* $c < 1$*.*

Finally, of particular relevance to the context of OpRisk modeling, the convergence of the partial sums can also be specified for a compound process setting according to the results in Theorem 7.5, see Sundt and Jewell (1981, theorem 1.11).

Theorem 7.5 *Consider the recursive relationship given by*

$$x_n = \left(a + \frac{b}{n}\right) x_{n-1}, \quad \forall n \geq s + 1, \tag{7.44}$$

with x_s*,* $a > 0$ *and* $b \geq -(s + 1)a$ *for some non-negative integer* s*. Then the infinite series* $\sum_{n=s}^{\infty} x_n$ *is convergent if* $a < 1$ *and divergent if* $a < 1$*. If* $a = 1$*, then it is convergent if* $b < -1$ *and divergent if* $b \geq -1$*.*

We finish this short section on properties of tail asymptotics for partial sums by mentioning two results on large deviations for partial sums. These two results provide bounds on the large deviations of partial sums which in turn provide information for studying the compound

(random sum) settings. We start with the result that follows from Cramer's theorem on large deviations which is given in Proposition 7.6.

Proposition 7.6 (Large Deviations for Partial Sums: Laplace Transform Conditions) *Consider an OpRisk loss model in which the severity distribution $X_i \sim F_X$ produces losses which are i.i.d. such that the Laplace transform $\mathbb{E}[\exp(\theta X_1)] < \infty$ is finite for some $t > 0$, for some sequence a_n that satisfies, as $n \to \infty$, the two conditions*

$$\frac{a_n}{\sqrt{n}} \to \infty, \quad \frac{a_n}{n} \to 0. \tag{7.45}$$

Then one can show that as the number of losses grows $n \to \infty$

$$\begin{aligned} \mathbb{P}\mathrm{r}\left[\sum_{i=1}^{n} X_i > n\mu + a_n\right] &= \exp\left(-\frac{1}{2\sigma^2}\frac{a_n^2}{n}\left[1 + o(1)\right]\right), \\ \mathbb{P}\mathrm{r}\left[\sum_{i=1}^{n} X_i < n\mu - a_n\right] &= \exp\left(-\frac{1}{2\sigma^2}\frac{a_n^2}{n}\left[1 + o(1)\right]\right), \end{aligned} \tag{7.46}$$

where $\mu = \mathbb{E}[X_i]$.

Then in Tang (2006) and Liu (2009), the following result in Proposition 7.7 for large deviations was developed which is based on moment conditions in the severity distribution satisfying some finite bound, in particular the existence of the mean. These results are, therefore, clearly related to the above-mentioned results in which restrictions were placed on the Laplace transform of the severity distribution.

Proposition 7.7 (Large Deviations for Partial Sums: Moment Conditions) *Consider an OpRisk loss model in which the severity distribution $X_i \sim F_X$ produces losses which are i.i.d. such that the r-th mean $\mathbb{E}[X_i^r]$ is finite for some $r > 1$. Then one can show that given $N = n$ losses have occurred, the partial sum satisfies the following tail bound, where for each $\alpha > 0$ and $\beta > 0$, there exists positive constants c_1 and $c_2 = C(c_1, \alpha)$ independent of x and the number of losses n such that when $x \geq \alpha n$, for $n = 1, 2, 3, \ldots$, one has*

$$\mathbb{P}\mathrm{r}\left[\sum_{i=1}^{n} X_i - n\mu > x\right] \leq n\mathbb{P}\mathrm{r}\left[X_1 - \mu > c_1 x\right] + c_2 x^{-\beta} \tag{7.47}$$

for $\mu = \mathbb{E}[X_i]$.

7.2 Asymptotics for LDA Models: Compound Processes

Once one allows the number of terms in the summation of the losses to be random, as in an LDA model compound process structure (random sum), then the study of the tail asymptotics of the resulting compound process loss distribution, that is, in OpRisk the annual loss distribution, will be affected by both the tails of the frequency and the tails of the severity distribution. This is a more delicate situation compared to the partial sum analysis discussed earlier. To clarify, now one considers the annual loss distribution given by the compound process

$$Z = \sum_{i=1}^{N} X_i \sim F_Z(z) = \sum_{i=0}^{\infty} \mathbb{P}\mathrm{r}[N = n] F_X^{(n)*}(x), \tag{7.48}$$

and in this section, the focus is on the tail features of $\overline{F}_Z$ as a function of $\overline{F}_X$ and $\mathbb{P}\mathrm{r}[N \geq n]$.

Perhaps the most straightforward setting involves the case in which the frequency distribution is light tailed and the severity distribution is light tailed; we only present a brief discussion later on this setting, introducing a basic saddlepoint series approximation that may be adopted for such settings.

Hence, one may consider the section of light-tailed frequency and severity as providing an important motivation that allows one to see which components of the study will be more complicated as the severity and frequency distribution become heavy tailed. By gaining an appreciation of this context, one can further understand the reason for more sophisticated approximations developed in the heavy-tailed settings. We start with the simplest possible way to introduce a saddlepoint series approximation based on the series approximation of each component of a truncated Poisson-mixture representation of the LDA model. Then we will present an alternative tail asymptotic expansion that is also based on a saddlepoint approximation developed in Embrechts *et al.* (1985a, p. 45–48) where they considered the behaviour of compound process tails comprised of severity distributions with light tails, as detailed in Theorem 7.6.

7.2.1 ASYMPTOTICS FOR LDA MODELS LIGHT FREQUENCY AND LIGHT SEVERITY TAILS: SADDLEPOINT TAIL APPROXIMATIONS

We will start by considering the setting in which the tails of the frequency and severity distributions are not heavy in the sense defined by the following two conditions:

$$\begin{aligned} &\mathbb{E}[\exp(\theta N)] < \infty \\ &\mathbb{E}[\exp(\theta X_i)] < \infty, \quad \forall i \in \{1, 2, 3, \ldots\}, \end{aligned} \tag{7.49}$$

that is, if one can show that the Laplace transform of the frequency and severity random variables is finite for some $\theta > 0$, then one can adopt classical asymptotic results for approximation of the annual loss models such as the saddlepoint approximation of Daniels (1954) and in the context of an LDA models, see Lugannani and Rice (1980). It was already discussed earlier how the bounding of the possible values that the Laplace transform can take directly relates to the existence of moments in the frequency and severity model. In this case, one can develop a special asymptotic tail approximation based on a series expansion, such as those studied previously in Cruz *et al.* (2015, chapters 17 and 18) for insurance modeling.

In this light-tailed case for the frequency and severity model, perhaps the simplest approach one may adopt to obtain an understanding of tail asymptotics of the annual loss distribution would be to develop an understanding of weighted combination of the tails for the severity given by truncating the number of terms in the summation to represent an upper bound on the possible number of losses and then to form an approximation of the tail asymptotics for each component partial sum distribution given by

$$\mathbb{P}\mathrm{r}[Z > z] = \sum_{i=0}^{\infty} \mathbb{P}\mathrm{r}[N = n] \overline{F_X^{(n)*}}(z)$$

$$\approx \frac{1}{\sum_{i=1}^{n_{\max}} \mathbb{P}\mathrm{r}[N=n]} \sum_{i=0}^{n_{\max}} \mathbb{P}\mathrm{r}[N=n]\overline{F_X^{(n)*}}(z). \tag{7.50}$$

Now approximating the tail asymptotics of the partial sum $\overline{F_X^{(n)*}}(z)$ could follow the results developed earlier if the severity distribution is heavy tailed, or alternatively, if the severity distribution is light tailed, then one can adopt a special class of series expansion for each partial sum distribution component in the mixture.

In developing a series expansion for the tail representation of the partial sum $\overline{F_X^{(n)*}}(z)$ distribution, we consider the approach developed in Lugannani and Rice (1980), Skovgaard (1987) and Daniels (1987). Lugannani and Rice (1980) considered the integral for the annual loss distribution given by integrating (inverse Fourier Transform) for the nth power of the characteristic function of the severity distribution given by

$$\begin{aligned} F_{Z_n}(z) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{\mathrm{i}\theta} \exp(-\mathrm{i}\theta x)[\Phi_X(\theta)]^n d\theta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{\mathrm{i}\theta} \exp\Big(n\left[\psi(\mathrm{i}\theta) - \mathrm{i}\theta\frac{z}{n}\right]\Big) d\theta, \end{aligned} \tag{7.51}$$

where $\exp(\psi(\mathrm{i}\theta)) = \Phi_X(\theta)$ and each component in the mixture for the partial sums of $\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > z | N = n]$ given by the asymptotic series approximation; note that the asymptotic here is not in n or in z but in the number of terms in the series representation that are considered

$$F_{Z_n}(z) \sim \frac{1}{2}\mathrm{erfc}\left(\sqrt{-f_0}\right) + \sum_{n=0}^{\infty} A_n - B_n \tag{7.52}$$

with the following terms definitions:

$$\begin{aligned} &f_0 = N\phi(\mathrm{i}\theta_0) - \mathrm{i}\theta_0 z, \\ &\frac{1}{2}\mathrm{erfc}(x) = \left(\frac{1}{2\pi}\right)^{-\frac{1}{2}} \int_{x\sqrt{2}}^{\infty} \exp\left(\frac{-t^2}{2}\right) dt, \\ &A_j = A_0 \frac{1}{n^j} \sum_{m=0}^{2j} (-\mu)^{2j-m} \sum_{l=0}^{m} d_{l,m}(-2)^{l+j} \left(\frac{1}{2}\right)_{l+j}, \quad \forall j \in \{1, 2, \ldots, n\}, \\ &d_{l+1,m+1} = \frac{1}{m+1} \sum_{k=1}^{m-l+1} k\eta_{k+1} d_{l,m-k+1}, \quad l \in [0, m], \\ &B_n = \frac{1}{2}(-\pi f_0)^{-\frac{1}{2}} f_0^{-n} \left(\frac{1}{2}\right)_n \exp(f_0), \end{aligned}$$

where the Pochhammer symbols are used with $(a)_n = a(a+1)(a+2)\cdots(a_n - 1)$. The other two components of this series representation that are important involve the initializations of the recursive evaluation for the coefficients A_j which involves the initializations

- if $d_{0,0} = 1$ and $d_{0,m} = 0$ for $m > 0$;
- if $l = 0$ and $m \geq 1$, then $d_{1,m} = \eta_{m+2}$; and
- if $l = m$, then $d_{m,m} = \frac{1}{m!}\eta_3^m$.

In Lugannani and Rice (1980), the first three series terms for A_j are provided:

$$A_0 = \frac{1}{\sqrt{2\pi n}}\mu \exp(f_0), \tag{7.53}$$

$$A_1 = -3A_0\frac{1}{n}\left[\frac{1}{3}\mu^2 + \mu\eta_3 + \frac{1}{2}\left(5\eta_3^2 - 2\eta_4\right)\right] \tag{7.54}$$

$$\begin{aligned} A_2 = 15A_0\frac{1}{n^2}\Bigg[&\frac{1}{5}\mu^4 + \mu^3\eta_3 + \frac{1}{2}\mu^2\left(7\eta_3^2 - 2\eta_4\right) \\ &+ \frac{1}{4}\mu(42\eta_3^3 - 28\eta_3\eta_4 + 4\eta_5) \\ &+ \frac{1}{8}\left(231\eta_3^4 - 252\eta_3^2\eta_4 + 56\eta_3\eta_5 + 28\eta_4^2 - 8\eta_6\right)\Bigg]. \end{aligned} \tag{7.55}$$

Here, the following definitions are utilized:

$$\begin{aligned} \psi^{(n)} &= \left[\mathrm{i}^{-n}\frac{d^n}{d\theta^n}\psi(\mathrm{i}\theta)\right]_{\theta=\theta_0}, \\ \eta_n &= \frac{\psi^{(n)}}{n!\left(\psi^{(2)}\right)^{\frac{n}{2}}}, \\ \mu &= \frac{1}{\mathrm{i}\theta_0\sqrt{\psi^{(2)}}}, \end{aligned}$$

with $\psi^{(n)}$ being the n-th cumulant of the severity distribution.

7.2.1.1 Existence of a Saddlepoint in Each Partial Sum. The above asymptotic saddlepoint series approximation relies on the existence and ability to be able to find the principle saddlepoint for each of the partial sums in the mixture for the annual loss distribution. To achieve this, one needs to first define the saddlepoint as the solution to a root search in the complex plane. The saddlepoint can be shown to be the solution θ_0 which is the principle saddlepoint of the equation $\exp(n\psi(\mathrm{i}\theta) - \mathrm{i}\theta z)$ and given as the solution to the equation

$$\frac{d}{d\theta}\psi(\mathrm{i}\theta) - \mathrm{i}\frac{1}{n}z = 0, \tag{7.56}$$

which, if it exists, will be located on the imaginary θ axis.

The existence of the saddlepoint was studied in Daniels (1954) where it was shown that θ_0 exists if for every value of $\frac{z}{N} \in [a, b]$ for some $a < b$ that may be infinite, the integral with respect to the severity distribution satisfies the condition

$$\mathbb{E}[\exp(\theta X)\mathbb{I}[x \in [a, b]] < \infty \tag{7.57}$$

for all real values of θ, see discussion in Lugannani and Rice (1980).

From these discussions, one can develop the tail representation for the LDA model based on the simplified saddlepoint series expansions given in Embrechts *et al.* (1985a, p. 45–48).

Theorem 7.6 (Compound Process Tails with Light-Tailed Severity Distributions) *Assume that the frequency distribution $p_n = \mathbb{P}\mathrm{r}[N = n]$ satisfies the condition that it is asymptotically equivalent for large n to a function parameterized by $w \in (0, 1)$, $\gamma \in \mathbb{R}$ and a function $C(n)$ which is slowly varying at ∞ according to*

$$p_n \sim w^n n^\gamma C(n),\ n \to \infty. \tag{7.58}$$

Consider the Laplace transform of the severity distribution given by

$$L_X(s) = \mathscr{L}[F_X(x)] = \int_0^\infty \exp(-sx)\, dF_X(x),\ \forall s \in \mathbb{R}. \tag{7.59}$$

Then, if the radius of convergence of the generating function of the frequency distribution satisfies the condition that there exists a positive number $\kappa > 0$ satisfying

$$w^{-1} = L_X(-\kappa) \tag{7.60}$$

for X non-arithmetic and such that $-L'_X(-\kappa) < \infty$, then the following asymptotic equivalence for the compound process tail distribution is satisfied,

$$\overline{F}_{Z_N}(x) \sim \frac{x^\gamma \exp(-\kappa x) C(x)}{\kappa(-wL'_X(-\kappa))^{\gamma+1}},\ \ x \to \infty. \tag{7.61}$$

The key message from this asymptotic result that is important to understand is that the right tail behaviour of the compound loss distribution is essentially determined by either the frequency or the severity components, depending on which has the heavier tail. This result cannot be applied in OpRisk in the standard Compound Poisson models as the tail of the frequency distribution is too light, resulting in the radius of convergence of the generating function being infinite. However, the result is applicable in OpRisk for numerous mixed Poisson distributions, including the negative binomial. In the case of the Poisson distribution and other frequency distributions not satisfying this radius of convergence requirements, one may make an alternative saddlepoint-based approximation, see an example which is detailed in Proposition 7.8. There are also numerous other conditions that can be developed for such approximations in the light-tailed cases.

Proposition 7.8 (Saddlepoint Approximations) *One can develop saddlepoint approximations for compound process tails with light-tailed bounded severity distributions in which the severity distribution is bounded with a bounded density, then the following saddlepoint approximation is applicable*

$$\overline{F}_{Z_N}(x) \sim \frac{\exp(\kappa x)}{|\kappa|(2\pi\lambda L''_X(\kappa))^{1/2}}[\exp(-\lambda(1 - L_X(\kappa)) - \exp(-\lambda)],\ \ x \to \infty, \tag{7.62}$$

where κ is the solution of $-\lambda L'_X(\kappa) = x$ and L_X is the Laplace transform of the severity distribution given by

$$L_X = \mathscr{L}[F(x)] = \int_0^\infty \exp(-sx)\, dF(x), \forall s \in \mathbb{R} \tag{7.63}$$

However, for many severity distributions of interest in OpRisk in the heavy-tailed setting, the Laplace transform does not exist, and as a consequence, results presented earlier do not apply. For example, if one considers the class of distributions where the tail decays like a power function, then they have the condition that their Laplace transform $L_X(-s) = \infty$ for all $s > 0$. However, this represents an important class of distributions in OpRisk which includes the Pareto, Burr, LogGamma, Cauchy, α-stable and t-distribution. Such distributions can be represented by a class of distribution functions with a regularly varying tail

$$\overline{F}_X(x) = x^\alpha C(x), \tag{7.64}$$

for $C(\cdot)$ some slowly varying function. It is in this context that we begin the discussions in this chapter relating to heavy-tailed severity distributions and appropriate approximations for the resulting single risk LDA compound process tails.

Next we consider the settings where the frequency, the severity or both components of the compound process annual loss distribution are heavy tailed in some sense as defined in Chapter 3. Each of the possible assumptions one may make regarding the tail heaviness of either the frequency or the severity model will impact on the asymptotic representations, as will be discussed throughout the remainder of this chapter.

7.3 Asymptotics for LDA Models Dominated by Frequency Distribution Tails

Once one allows the number of terms in the summation of the losses to be random, as in an LDA model compound process structure (random sum), then the study of the tail asymptotics of the resulting compound process loss distribution, that is, in OpRisk, the annual loss distribution, will be affected by both the tails of the frequency and the tails of the severity distribution. This is a more delicate situation compared to the partial sum analysis discussed earlier. To clarify, now one considers the annual loss distribution given by the compound process

$$Z = \sum_{i=1}^{N} X_i \sim F_Z(z) = \sum_{i=0}^{\infty} \mathbb{P}\mathrm{r}[N = n] F_X^{(n)*}(x), \tag{7.65}$$

and in this section, the focus is on the tail features of $\overline{F}_Z$ as a function of $\overline{F}_X$ and $\mathbb{P}\mathrm{r}[N \geq n]$. In particular, we will first be interested in the case in which the frequency distribution tails dominate the tail behaviour of the severity model.

Going forward, we will consider the ideas of Stam (1973) and more recently in Robert and Segers (2008), where two cases are considered to classify the possible outcomes. If the frequency distribution is regularly varying with finite mean and the severity distribution is regularly varying with finite mean, then one has one of the two outcomes for $\alpha > 0$ and $\beta > 0$ given for $\gamma > 0$ by

$$\lim_{x \to \infty} \frac{\mathbb{P}\mathrm{r}[N > x\gamma]}{\mathbb{P}\mathrm{r}[N > x]} = \gamma^{-\alpha} \tag{7.66}$$

and

$$\lim_{x \to \infty} \frac{\mathbb{P}\mathrm{r}[X_1 > x\gamma]}{\mathbb{P}\mathrm{r}[X_1 > x]} = \gamma^{-\beta}. \tag{7.67}$$

Under either of these above assumptions and, in addition, if X_i is independent of N for all losses, then one of the two conditions can be shown to arise for the resulting tail behaviour of the annual loss as $z \to \infty$ according to

$$\begin{aligned} &\mathbb{P}\mathrm{r}[Z > z] \sim \mathbb{E}[N]\mathbb{P}\mathrm{r}[X_1 > z], \quad \textit{if} \quad \alpha > \beta \\ &\mathbb{P}\mathrm{r}[Z > z] \sim \mathbb{P}\mathrm{r}[\mathbb{E}[X_1]N > z], \quad \textit{if} \quad \beta > \alpha. \end{aligned} \tag{7.68}$$

The first is the single-loss approximation (severity tails dominate), and this will be studied in future sections of this chapter. In this first case, the severity distribution is sub-exponential but the frequency distribution is lighter tailed; then one can also adopt the representation known as the first-order approximation that will be presented later. Here we first focus on the second case where the frequency distributions tails dominate the tails of the severity distribution.

In this section, the primary concern involves the understanding of what happens to an LDA model's asymptotic tail behaviour when the tail of the compound process Z is dominated by the tail of the frequency distribution, the so-called heavy-tail frequency, light-tail severity models. In Rolski *et al.* (2009) and Robert and Segers (2008), they studied the question related to when does the following asymptotic equivalence condition arise in an LDA model structure:

$$\overline{F}_Z(z) = \mathbb{P}\mathrm{r}[Z > z] \sim \mathbb{P}\mathrm{r}[\mathbb{E}[X_1]N > z], \quad z \to \infty. \tag{7.69}$$

Remark 7.4 *It will be shown that one can address this question of validity of this asymptotic equivalence even in the general case where the number of losses N and the severity loss amounts X_i are not independent.*

To proceed, the frequency distribution will be assumed to be in the maximal domain of attraction of an extreme value distribution, which must be either Frechet or Gumbel because the frequency distributions take discrete support. This has implications for the heavy-tailed features of the frequency distributions being considered, in the following results as depending on which of the resulting classes of EVT distribution, Frechet and Gumbel, is considered will affect the tail representation considered for the heavy-tailed frequency distribution. For instance, it turns out that, as discussed in Stam (1973), Faÿ *et al.* (2006) and Robert and Segers (2008), one can characterize the consistently varying (CV) frequency distributions, those that satisfy the condition given in Definition 3.19, by considering the frequency distributions and their domain of attraction to EVT families. This will be useful as the proceeding results in this subsection will rely on the properties of consistent variation of the frequency distribution.

Proposition 7.9 (Consistent Variation and Frechet Domain of Attraction) *A frequency distribution function $F_N(n)$ can be shown to be in the class of CV distributions $F_N \in CV$ by showing it is in the class of regularly varying functions, where $RV \subset CV$, which have representation in terms of the slowly varying function $L(n)$ given by*

$$\overline{F}_N(n) = n^{-\alpha} L(n). \tag{7.70}$$

If the frequency distribution satisfies the sufficient condition given by the Von Mises-type condition:

$$\lim_{n\to\infty} \frac{n\mathbb{P}\mathrm{r}[N = n]}{\mathbb{P}\mathrm{r}[N > n]} = \alpha, \tag{7.71}$$

then it will be represented as in Equation 7.70 and will be, therefore, in the class of regularly varying functions. This also means that the distribution for the frequency N will necessarily be in the domain of attraction of Frechet distribution with shape parameter $\alpha > 0$.

As in Robert and Segers (2008), the frequency distribution will be a member of the maximal domain of attraction of the Gumbel EVT model if the following asymptotic representation in Proposition 7.10 holds; this is a repeat of the results found in Chapter 2 for convenience.

Proposition 7.10 (Gumbel Domain of Attraction for Heavy-Tailed Frequency) *A frequency distribution $F_N(n)$ is in the maximal domain of attraction of the Gumbel EVT model if it has an asymptotic representation given by*

$$\lim_{n\to\infty} \frac{\mathbb{P}\mathrm{r}[N > n + \gamma a(n)]}{\mathbb{P}\mathrm{r}[N > n]} = \exp(-\gamma), \quad \gamma \in \mathbb{R}, \tag{7.72}$$

for a function $a(n)$ *which is asymptotically equivalent to the mean excess function of the frequency distribution as* $n \to \infty$

$$a(n) \sim \mathbb{E}[N - n | N > n] \tag{7.73}$$

and also satisfies the conditions

$$\begin{aligned} \lim_{n\to\infty} \frac{a(n)}{n} &= 0, \\ \lim_{n\to\infty} \frac{a(n + \gamma a(n))}{a(n)} &= 1. \end{aligned} \tag{7.74}$$

If the frequency distribution satisfies these conditions, it will have a tail representation given by

$$\overline{F}_N(n) = c(n) \exp\left(- \int_0^n \frac{g(t)}{a(t)} dt \right) \tag{7.75}$$

with $\lim_{n\to\infty} c(n) = c \in (0, \infty)$ *and* $\lim_{t\to\infty} g(t) = 1$.

7.3.1 HEAVY-TAILED FREQUENCY DISTRIBUTION AND LDA TAIL ASYMPTOTICS (FRECHET DOMAIN OF ATTRACTION)

Given these precise definitions just presented above for what we mean when we refer in this section to the heavy-tailed frequency distribution and the resulting representations such definitions provide, we now consider the case in which the frequency distribution for the number of losses in the compound process will be of consistent variation $F_N \in CV$. We consider the results developed in Robert and Segers (2008) which provide conditions for the validity of the tail asymptotic equivalence between the compound process annual loss tails and the frequency distribution tails given in Equation 7.69, see Proposition 7.11.

Proposition 7.11 (Heavy-Tailed Frequency and LDA Tail: Dependent Frequency *N* and Severity *X* Case) *Consider an OpRisk LDA loss model in which the frequency distribution* $N \sim F_N$ *satisfies that* $F_N \in CV$ *where it is of consistent variation and, therefore, satisfies*

$$\lim_{\gamma\uparrow 1} \limsup_{x\to\infty} \frac{\overline{F}_N(x\gamma)}{\overline{F}_N(x)} = 1, \tag{7.76}$$

and the severity distribution satisfies, for all losses, two conditions

1. *finite moments of the power* $\mathbb{E}[X_i^r] < \infty$ *for some* $r > 1$*; and*
2. *the tail asymptotic condition*

$$x\mathbb{P}\mathrm{r}[X_1 > x] = o(\mathbb{P}\mathrm{r}[N > x]). \tag{7.77}$$

Then one can show the following annual loss asymptotic, as $x \to \infty$*:*

$$\overline{F}_Z(x) = \mathbb{P}\mathrm{r}[Z > x] \sim \mathbb{P}\mathrm{r}\left[N > \frac{x}{\mu} \right]. \tag{7.78}$$

Remark 7.5 *The above asymptotic LDA tail property will hold for arbitrary-dependent structures between the frequency and severity distributions in the OpRisk LDA model, see discussions in Robert and Segers (2008).*

This result can be further refined as shown in Robert and Segers (2008) if one assumes independence between the frequency and severity loss random variables in the LDA model, as detailed in Proposition 7.12.

Proposition 7.12 (Heavy-Tailed Frequency and LDA Tail: Independent Frequency *N* and Severity *X* Case) *Consider an OpRisk LDA loss model in which the frequency distribution* $N \sim F_N$ *satisfies that* $F_N \in CV$ *where it is of consistent variation and, therefore, satisfies*

$$\lim_{\gamma \uparrow 1} \limsup_{x \to \infty} \frac{\overline{F}_N(x\gamma)}{\overline{F}_N(x)} = 1, \tag{7.79}$$

and independent of N*, one has the i.i.d. loss random variables* $\{X_i\}_{i=1}^N$ *that satisfy the moment condition for some* $r > 1$

$$\mathbb{E}[X_1^r] < \infty, \tag{7.80}$$

then one can show under one of the two cases depending on the existence of the frequency distribution mean, given by:

1. *Case 1. If the mean of the severity distribution is finite* $\mathbb{E}[N] < \infty$ *and*

$$\mathbb{P}\mathrm{r}[X_1 > x] = o(\mathbb{P}\mathrm{r}[N > x]), \quad x \to \infty, \textit{or alternatively}; \tag{7.81}$$

2. *Case 2. If the mean of the frequency distribution is infinite* $\mathbb{E}[N] = \infty$ *and there exists a constant* $c \in [1, r)$ *such that*

$$\limsup_{x \to \infty} \frac{\mathbb{E}[N\mathbb{I}[N \leq x]]}{x^q \mathbb{P}\mathrm{r}[N > x]} < \infty, \tag{7.82}$$

that the following compound process asymptotic result is obtained

$$\overline{F}_Z(z) = \mathbb{P}\mathrm{r}[Z > x] \sim \mathbb{P}\mathrm{r}\left[N > \frac{x}{\mu}\right], \quad x \to \infty \tag{7.83}$$

with $\mu = \mathbb{E}[X_1]$.

7.3.2 HEAVY-TAILED FREQUENCY DISTRIBUTION AND LDA TAIL ASYMPTOTICS (GUMBEL DOMAIN OF ATTRACTION)

Throughout this section, it is assumed that the severity distribution has a moment-generating function that is finite in the neighbourhood of the origin. Furthermore, the number of losses in a year is modeled by a heavy-tailed frequency distribution in the domain of attraction of the Gumbel EVT model, as discussed at the start of this section, which means it satisfies the limiting condition

$$\lim_{n \to \infty} \frac{\mathbb{P}\mathrm{r}[N > n + \gamma a(n)]}{\mathbb{P}\mathrm{r}[N > n]} = \exp(-\gamma), \quad \gamma \in \mathbb{R}. \tag{7.84}$$

Recall that the function $a(n)$ which is asymptotically equivalent to the mean excess function of the frequency distribution as $n \to \infty$

$$a(n) \sim \mathbb{E}[N - n | N > n] \tag{7.85}$$

and satisfies the conditions

$$\lim_{n\to\infty} \frac{a(n)}{n} = 0,$$
$$\lim_{n\to\infty} \frac{a(n + \gamma a(n))}{a(n)} = 1. \tag{7.86}$$

Hence, if the mean excess function is asymptotically satisfying the condition

$$\lim_{n\to\infty} \frac{a(n)}{n^{2/3}} \to \infty, \tag{7.87}$$

then it was shown in Robert and Segers (2008) that one will obtain the result for the annual loss asymptotic given by

$$\overline{F}_Z(z) = \mathbb{P}\mathrm{r}[Z > z] \sim \mathbb{P}\mathrm{r}[\mathbb{E}[X_1]N > z], \quad z \to \infty. \tag{7.88}$$

In other words, one has the asymptotic equivalence between the tails of the annual loss severity distribution and the tail of the frequency distribution, given by:

$$\overline{F}_Z(z) \sim \overline{F}_N(z/\mu). \tag{7.89}$$

This result was further refined in the case that the number of losses N and the loss amounts X_i are independent as shown in Proposition 7.13, see Robert and Segers (2008).

Proposition 7.13 *Consider an OpRisk LDA loss model in which the frequency distribution $N \sim F_N$ satisfies that F_N is in the maximal domain of attraction of the Gumbel distribution. Furthermore, assume that N is independent of the severity amounts X_i for all losses $i \in \{1, 2, \ldots, N\}$ and that the severity distribution satisfies that the transform $\mathbb{E}[\exp(\gamma X_1)] < \infty$ for a constant $\gamma > 0$ and a function $a(n)$ satisfying*

$$\liminf_{n\to\infty} \frac{\ln a(n)}{\ln n} > \frac{1}{2}, \tag{7.90}$$

then one obtains the annual loss asymptotic equivalence given by

$$\overline{F}_Z(z) = \mathbb{P}\mathrm{r}[Z > z] \sim \mathbb{P}\mathrm{r}[\mathbb{E}[X_1]N > z], \quad z \to \infty; \tag{7.91}$$

in other words, one has the asymptotic equivalence $\overline{F}_Z(z) \sim \overline{F}_N\left(\frac{z}{\mu}\right)$.

Hence, these sets of results characterize the conditions under which the frequency distribution dominates the severity distribution tails and the resulting tail asymptotics for the annual loss that can be derived as a consequence. In the following sections, we consider the settings in which the severity distribution is dominating in tail behaviour over the frequency distribution.

7.4 First-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses

As has been discussed earlier, the saddlepoint-type approximations are only suitable under certain properties of the frequency and severity model in the LDA structure. However, for many severity distributions of interest in OpRisk in the heavy-tailed setting, the Laplace transform

does not exist and as a consequence, results presented earlier do not apply. For example, if one considers the class of distributions where the tail decays like a power function, then they have the condition that their Laplace transform $L_X(-s) = \infty$ for all $s > 0$. However, this represents an important class of distributions in OpRisk which includes the Pareto, Burr, LogGamma, Cauchy, α-stable and t-distribution. Such distributions can be represented by a class of distribution functions with a regularly varying tail

$$\overline{F}_X(x) = x^{\alpha} C(x), \tag{7.92}$$

for $C(\cdot)$ slowly varying. It is in this context that we begin the discussions in this chapter relating to heavy-tailed severity distributions and appropriate approximations for the resulting single risk LDA compound process tails. This leads one to consider alternative asymptotic representations of the LDA annual loss distribution tail, the simplest of these when the severity distribution is heavy tailed is known as the first-order approximation.

To understand the derivation of first-order single risk loss process right tail asymptotics, we consider the approach discussed in Daley *et al.* (2007). If one considers the compound process distribution and right tail distribution functions given, for annual loss $Z_N = \sum_{i=1}^{N} X_i$, given by

$$F_{Z_N}(x) = \sum_{n=0}^{\infty} p_n F_X^{(n)*}(x)$$

$$\overline{F_{Z_N}}(x) = \sum_{n=1}^{\infty} p_n \overline{F_X^{(n)*}}(x),$$

where it is assumed that the frequency probability mass function satisfies $\sum_{n=0}^{\infty} np_n < \infty$. An understanding of the asymptotic approximation of the annual loss distribution in each case is derived from finding an upper bound on the ratio of the partial sum tail distribution and the severity distribution tail which will hold for all $n \in \mathbb{J}^+$ such that

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq g(n), \quad \forall x \geq 0. \tag{7.93}$$

The form of the bounding function $g(n)$ will be seen to depend on the assumptions one makes about the tail behaviour of the frequency and severity models. We start this section with a general assumption that the severity distribution is sub-exponential, and then we step through how to obtain the first-order asymptotic for the resulting annual loss distribution that naturally follows from such a bound. This is then progressively refined in the first-order case via different assumptions.

7.4.1 FIRST-ORDER SINGLE RISK LOSS PROCESS ASYMPTOTICS FOR HEAVY-TAILED LDA MODELS: GENERAL SUB-EXPONENTIAL SEVERITY MODEL RESULTS

The classical first-order SLA as discussed in Albrecher *et al.* (2010) takes the form given by Definition 7.5. To understand the first-order asymptotic behaviour of the right tail $\overline{F_{Z_N}}(x)$, one can consider obtaining an upper bound for the asymptotic ratio of the tail of the compound distribution and the severity distribution tail for each number of losses $n \in \mathbb{J}$. This is given in Lemma 7.5 for the standard geometric Kesten bound and then, in the more general bound, in Theorem 7.8 derived from Daley *et al.* (2007, theorem 7).

Lemma 7.5 (Kesten's Bound) *If the severity distribution F is in the class of sub-exponential distributions, then for each $\varepsilon > 0$, there exists a constant $K = K(\varepsilon) < \infty$ such that for all $n \geq 2$, the following bound holds*

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq K(1+\varepsilon)^n, \ \ \forall x \geq 0. \tag{7.94}$$

Given the Kesten bound, one now applies the dominated convergence theorem given in Lemma 7.6 to the limit in Equation 7.99 to obtain the result for the first-order single-loss approximation in Definition 7.5.

Lemma 7.6 (Dominated Convergence Theorem) *Consider a sequence of integrable functions $\{f_n\}$ on probability space (Ω, Σ, μ) which satisfy the limit that*

$$\lim_{n\to\infty} f_n(x) = f(x), \text{ almost everywhere} - \mu. \tag{7.95}$$

In addition, suppose that there exists an integrable function

$$H \geq 0 : |f_n(x)| \leq H(x), \ \forall n.$$

Then if $|f(x)| \leq H(x)$, one has the limit result

$$\lim_{n\to\infty} \int_\Omega f_n(x) d\mu(x) = \int_\Omega f(x) d\mu(x). \tag{7.96}$$

Therefore, one utilize this bound and the dominated convergence theorem to interchange the order of the summation and the limit and then utilize the fact that for heavy-tailed sub-exponential severity models, the condition that

$$\lim_{x\to\infty} \frac{\overline{F_X^{(2)*}}(x)}{\overline{F}_X(x)} = 2 \tag{7.97}$$

characterizing this sub-class of distributions implies for such subexponential severity models that

$$\lim_{x\to\infty} \frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} = n, \tag{7.98}$$

see Embrechts *et al.* (1997, lemma 1.3.4). Hence, one obtains for the compound process ratio,

$$\lim_{x\to\infty} \frac{\overline{F}_{Z_N}(x)}{\overline{F}_X(x)} = \lim_{x\to\infty} \sum_{n=1}^{\infty} p_n \frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} = \sum_{n=1}^{\infty} n p_n = \mathbb{E}[N], \tag{7.99}$$

which is equivalent to the asymptotic equivalence statement that

$$\overline{F}_{Z_N}(x) \sim \mathbb{E}[N]\overline{F}_X(x). \tag{7.100}$$

This can then be utilized to state the first-order result in Definition 7.5.

Definition 7.5 (First-Order Single-Loss Approximation of Compound Process Tail) *Given a sequence of i.i.d. losses X_i with sub-exponential severity distribution function F_X and an independent integer-valued random variable for the number of losses, N satisfying*

$\mathbb{E}[(1+\varepsilon)^N] < \infty$, *for an* $\varepsilon > 0$. *Then the first-order asymptotic approximation of the compound process tail probabilities, as* $x \to \infty$, *is given by*

$$\Pr[X_1 + X_2 + \cdots + X_N \geq x] = \overline{F}_Z(x) = a_1(x) + o(\overline{F}_X(x)) \tag{7.101}$$

with

$$a_1(x) = \mathbb{E}[N]\overline{F}_X(x). \tag{7.102}$$

This can be restated analogously for the annual loss tail as defined in Definition 2.3 for situations in which we consider sub-exponential family members for the severity distribution, $F_X(x) \in \mathscr{F}$. In particular, the result in Theorem 7.7, from Embrechts *et al.* (1997, theorem 1.3.9), states that in situations in which one may assume that the probability-generating function for the frequency random variable N is analytic at $z = 1$, then the tail of the annual loss distribution is well approximated by the tail of the maximum of the i.i.d. sequence of losses, $X_1, \ldots, X_n$. This is detailed by Theorem 7.7 detailing the first version of an SLA.

Theorem 7.7 (Asymptotic Tail Approximation for Single Risk Sub-exponential Severity) *Consider an annual loss distribution* $G(z) = F_Z(z)$ *under an LDA formulation,*

$$\begin{aligned} G(z) = F_Z(z) &= \Pr[Z \leq z] \\ &= \sum_{n=0}^{\infty} \Pr[Z \leq z | N = n] \Pr[N = n] \qquad (7.103) \\ &= \sum_{n=0}^{\infty} p_n F_X^{(n)*}(z), \qquad (7.104) \end{aligned}$$

with the severity distribution satisfying $f_X(z) \in \mathscr{F}$. *Furthermore, assuming that for some* $\varepsilon > 0$,

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n p_n < \infty. \tag{7.105}$$

Then the right tail of the annual loss distribution $F_Z(z)$ *for the annual loss random variable* Z *is approximated according to an SLA given by*

$$\overline{F}_Z(x) = \mathbb{E}[N]\overline{F}_X(x)(1 + o(1)) \text{ as } x \to \infty \tag{7.106}$$

or, equivalently, the tail of the annual loss distribution is sub-exponential in behaviour with asymptotic equivalence,

$$\overline{F}_Z(x) \sim \mathbb{E}[N]\overline{F}_X(x).$$

This result makes intuitive sense, because in the heavy-tailed sub-exponential case by the tail of the distribution of the sum and the maximum are asymptotically of the same order. Hence, the sum of losses is governed by large individual losses, and, therefore, the distributional tail of the sum is mainly determined by the tail of the maximum loss.

To conclude the first-order asymptotic results for the sub-exponential case, consider the Poisson compound process with severity distribution satisfying that it is locally sub-exponential, where each loss is i.i.d. with $X_i \sim F_{X_i} \in \mathscr{F}_\Delta$ for some $T \leq \infty$ such that $\Delta = [0, T)$, then an analogous first-order asymptotic result can be obtained as shown in Proposition 7.14, see Wang *et al.* (2005).

Proposition 7.14 (Asymptotic Tail Approximation with Locally Sub-exponential Severity) *Consider compound process LDA model with frequency distribution given by a Poisson distribution $N \sim Poisson(\lambda)$ and loss random variables which are i.i.d. with loss distribution f_X which satisfies that it is locally sub-exponential such that $F_X \in \mathcal{F}_\Delta$ for some $T \leq \infty$ such that $\Delta = [0, T)$. Then the first-order asymptotic is given by*

$$F_{Z_N}(x + \Delta) \sim \lambda F_X(x + \Delta). \tag{7.107}$$

Clearly, if $T = \infty$, then one recovers the standard result for first-order asymptotics of the LDA model with sub-exponential severity distributions.

Remark 7.6 *We note that the results presented in Definition 7.5 of the first-order single-loss tail approximation for sub-exponential models required that $\mathbb{E}[z^N] < \infty$ for some $z > 1$ which means that N must have finite moments of all orders. This clearly has implications on the properties of the severity distribution because we observed the result that*

$$\lim_{x \to \infty} \frac{\overline{F}_{Z_N}(x)}{\overline{F}_X(x)} = \mathbb{E}[N]. \tag{7.108}$$

The restrictions required on the frequency distribution for the first-order asymptotic result to hold can be relaxed. This is achieved by considering what results presented earlier for the whole sub-exponential family can be achieved for either the sub-family of the sub-exponential family such as severity models with right tails that are of regular variation or, alternatively, asking what can be said about the larger class of tail function behaviours as characterized by the O-regular variation family and their intersection with other families of tail behaviours such as long-tailed distributions.

7.4.2 FIRST-ORDER SINGLE RISK LOSS PROCESS ASYMPTOTICS FOR HEAVY-TAILED LDA MODELS: REGULAR AND O-REGULARLY VARYING SEVERITY MODEL RESULTS

In deriving the first-order tail approximations for the family of sub-exponential severity models, the significance of the role played by bounding the tail ratios of the partial sum distribution and the severity distribution was clear. So we begin this section by considering how the Kesten bound may be generalized to settings in which the severity distribution considered is either $F_X \in RV$ or $F_X \in OR$. In such cases, it can be shown that one of the two following results may be achieved for the function $g(n)$ providing the upper bound

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq g(n), \quad \forall x \geq 0, \tag{7.109}$$

depending on the assumptions on the severity model as follows (see Daley *et al.* (2007)):

1. If $F_X \in RV(-\alpha)$ for any $\alpha > 0$, then one can show that $g(n)$ is given by

$$g(n) = An^{\alpha+1+\varepsilon} \tag{7.110}$$

for some constant A that is independent of the severity x and the number of terms in the partial sum n.

2. If $\overline{F}_X \in OR$, then one can obtain a result for $g(n)$ of the same form as above with α replaced by the lower Matuszewska index $\beta = \beta(\overline{F}_X)$, such that for $\varepsilon > 0$, one can find a positive constant A that will satisfy

$$g(n) = An^{|\beta|+1+\varepsilon}, \ \ \forall x \geq 0. \tag{7.111}$$

Note: Recall the following result to see how this assumption differs from the assumption that F_X is sub-exponential. If $\overline{F}_X$ is in the class intersection of O-regularly varying functions and long-tailed functions $\overline{F}_X \in OR \cap \mathscr{L}$, then this implies that the severity distribution is sub-exponential $F_X \in \mathscr{F}$.

3. If $F_X \in RV_{-\alpha}$ for $\alpha \in (0, 1)$, then one can show that $g(n) = An$.

4. If $F_X \in OR$ and the lower Matuszewska index $\beta(\overline{F}_X) > -1$, then $g(n) = An$.

The general upper bound for the tail ratios for each number of losses n of $\overline{F_X^{(n)*}}(x)/\overline{F}_X(x)$ is given in Theorem 7.8, see Daley *et al.* (2007).

Theorem 7.8 (Generalized Tail Ratio Bounds for Partial Sums and Severity Models) *If one considers a severity distribution F_X for which the only assumption is that it is from the family of sub-exponential distributions. Next define the sequence of constants $\{\alpha_n\}_{n\geq 0}$ for supremum given by*

$$\alpha_n = \sup_{x\geq 0} \frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)}, \tag{7.112}$$

then one obtains a general bound on α_n for each n which is only a function of $\overline{F}_X$ and $\overline{F_X^{(2)*}}$ using for any $n = 1, 2, \ldots$

$$\alpha_n \leq \sum_{j=1}^{n-1}(1 + c_j) \prod_{k=j+1}^{n-1}(1 + \varepsilon_k) + \prod_{k=1}^{n-1}(1 + \varepsilon_k) \tag{7.113}$$

with the empty product $\prod_{k=1}^{n-1}(1 + \varepsilon_k) = 1$ and where the sequence T_n, ε_n are defined to satisfy

$$\sup_{x\geq T_n} \frac{\overline{F_X^{(2)*}}(x)}{\overline{F}_X(x)} - 1 \leq 1 + \varepsilon_n, \tag{7.114}$$

where ε_n can be selected arbitrarily small through selection of T_n large and c_n is given by

$$c_n = \frac{F_X(T_n)}{\overline{F}_X(T_n)}. \tag{7.115}$$

Remark 7.7 *To understand how this bound would behave if one assumed that F_X was sub-exponential and, therefore, how this bound is compared to the Kesten bound, consider the case that $\overline{F_X^{(2)*}}(x)/\overline{F}_X(x)$ is monotonically increasing beyond some point for $x \geq x_0$ and select the value $T_n = x_0$ with $\varepsilon_n = 0$, and one will achieve a constant*

$$\alpha_n \leq \sum_{j=1}^{n-1}(n-1)\left(1 + \frac{F_X(x_0)}{\overline{F}_X(x_0)}\right) + 1. \tag{7.116}$$

If x_0, then one has $\alpha_n \leq n$.

Next we preset a basic discussion on the context of the development of these bounds above, which, as in the case of the Kesten bound used for sub-exponential distributions, allows one to obtain the first-order approximations of the compound process tails. Generalizations of the results for first-order asymptotics from the sub-exponential case were first studied based on the aspects of the above bound functions $g(n)$ in the work of Stam (1973). In this work, they considered the first-order tail asymptotics under the context of regular variation of the severity model. In particular, they considered under what conditions can the first-order asymptotic result

$$\overline{F}_Z(x) \sim \mathbb{E}[N]\overline{F}_X(x) \tag{7.117}$$

still hold if one assumes a more restrictive sub-class of severity model tail properties given by the notion of regular variation. They considered in which cases one can relax other assumptions on the LDA model as a consequence. The assumption was that the severity distribution tail satisfies $\overline{F}_X \in RV(-\alpha)$, that is, one can represent the severity distribution tail via a power law and a slowly varying function $L(x) \in RV(0)$ according to

$$\overline{F}_X(x) = x^{-\alpha}L(x). \tag{7.118}$$

In order to obtain the bounds required to satisfy the first-order asymptotic result, one may pose the problem slightly more generally, for instance, in considering the ratio of the tail behaviour of the partial sums and severity distribution tails given by

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq g(n), \quad \forall x \geq 0, \tag{7.119}$$

one may instead consider the analogous relationships between a generic regularly varying function $V(x) \in RV$ and the severity distribution tail $\overline{F}_X$ under different tail variation assumptions; the compound process tail $\overline{G} = \overline{F}_{Z_N}$, that is, each of the partial sums for all $n \geq 1$ under different tail variation assumptions; and the frequency distribution under different tail variation assumptions. Such a study will result analogously in the bounds discussed earlier and, therefore, allow one to derive first-order asymptotic results.

The three different tail asymptotic cases were studies as $x \to \infty$ for some regularly varying function $V(x) \in RV_\rho$ and the severity distribution f_X or annual loss distribution $G = F_{Z_N}$ and constants a, b, c (see Stam 1973 and later in Daley *et al.* 2007):

Case 1. A regularly varying function $V(x)$ and the tail of the severity distribution which was assumed to be regularly varying which will satisfy $V(x)\overline{F}_X(x) \to a \geq 0$;

Case 2. A regularly varying function $V(x)$ and the tail of the annual loss distribution when the severity distribution tail was assumed to be regularly varying which will satisfy $V(x)\overline{F}_X(x) \to b$; and

Case 3. A regularly varying function $V(x)$ and the tail of the frequency distribution which will satisfy $V(x)\mathbb{P}\mathrm{r}[N > x] \to c \geq 0$.

The question of under what conditions will such asymptotic convergence hold was studied, that is, the question of when can one obtain bounds for the tail ratios discussed earlier, was addressed based on Cases 1–3.

7.4.2.1 Conditions for Convergence of Cases 1–3. The answers to this question of validity of Cases 1–3 were provided first in some contexts in Stam (1973) and then generalized by Daley *et al.* (2007) where they stated the following results:

1. If the index of regular variation of the function $V(x) \in RV_\rho$ is such that $\rho > 1$, then one can show that Case 1 will hold iff Case 2 holds with the constant given by

$$b = c\mathbb{E}[X]^\rho + a\mathbb{E}[N], \tag{7.120}$$

which implies a finite mean for the severity model. To relate this back to the first-order tail approximation of the annual loss distribution, one observes that if for the given function $V(x)$ either of the following two conditions applies:

- the tail of the frequency distribution also satisfies that

$$V(n)\mathbb{P}\mathrm{r}[N > n] \to 0 \;\; as \;\; n \to \infty; \;\; \text{or} \tag{7.121}$$

- alternatively, if $\alpha > 1$ for the tail regular variation of the severity model and one selects $V(x) = \frac{1}{\overline{F}_X(x)} \in RV(\alpha)$ with $a = 1$, then given that $\mathbb{E}[N^{\alpha+\varepsilon}] < \infty$, then one can show that

$$n^{\alpha+\varepsilon}\mathbb{P}\mathrm{r}[N > n] \to 0 \;\; as \;\; n \to \infty, \tag{7.122}$$

then one has the desired first-order asymptotic holding

$$\overline{F}_Z(x) \sim \mathbb{E}[N]\overline{F}_X(x). \tag{7.123}$$

2. If the index of regular variation of the function is $V(x) \in RV(\rho)$ such that $\rho \in [0, 1)$ and one has finite mean for the frequency distribution such that $\mathbb{E}[N] < \infty$, then one can show that Case 1 holds iff Case 2 holds with constant given by

$$b = a\mathbb{E}[N], \tag{7.124}$$

which implies that the mean for the severity model is not finite.

3. If the index of regular variation of the function is $V(x) \in RV(\rho)$ such that $\rho = 1$, this special case was studied in Stam (1973), Embrechts *et al.* (1985b) and Mallor and Omey (2006).

The assumptions required for the existence of a first-order asymptotic result for the annual loss tail approximation were then refined by considering the class of severity model tail behaviours given by the O-regularly varying severity distribution tails. This study was undertaken in Daley *et al.* (2007) where they weaken the condition of all moments of N being finite, based on the results presented earlier from Stam (1973).

This relaxation involves considering less restrictive tail assumptions on the sub-exponential severity distribution $\overline{F}_X$, by restricting the possible severity model characteristics to a sub-family of the sub-exponential family; in particular, it is assumed instead that it is a function of O-regular variation $\overline{F}_X \in OR$. Then two bounds are obtained for the ratio $\overline{F_X^{(n)*}}(x)/\overline{F}_X(x)$ depending on whether the lower Matuszewska index $\beta(\overline{F}_X)$ is in the interval $(-\infty, -1]$ or $(-1, 0)$; then the corresponding bounds can be obtained where

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq An^{|\beta(\overline{F})|+1+\varepsilon}, \text{ for } \overline{F}_X \in OR \text{ and } \beta(\overline{F}_X) < -1 \tag{7.125}$$

for some constants A and $\varepsilon > 0$ or

$$\frac{\overline{F_X^{(n)*}}(x)}{\overline{F}_X(x)} \leq An, \text{ for } \overline{F}_X \in OR \text{ and } \beta(\overline{F}_X) \in (-1, 0) \tag{7.126}$$

for some constant A. Then again, given these tail ratio bounds, one may obtain the existence of the first-order tail asymptotic as presented.

A summary of their results is provided in Daley *et al.* (2007, theorem 2) as shown in Theorem 7.9 which provides a detailed set of conditions one can assume in order that the tail ratio of the n-fold convolution of the severity distribution and the severity distribution is bounded and, therefore, that the limiting asymptotic tail ratio of the compound distribution of the annual loss over the tail behaviour of the severity distribution is finite.

Theorem 7.9 (Asymptotic Tail Approximations for Single Risk with RV or OR Severity) *Consider the compound process*

$$F_{Z_N}(x) = \sum_{n=1}^{\infty} p_n F_X^{(n)*}(x)$$

with sub-exponential severity distribution F_X and frequency distribution specified by $\{p_n : n \in \mathbb{J}\}$. Then one can show the following results relating to the ratio of tail of the compound process to the tail of the severity distribution and the bounds on this ratio for a given n take one of the following two forms:

1. *If $\sum_{n=1}^{\infty} np_n = \infty$, then $\liminf_{x\to\infty} \frac{\overline{F_{Z_N}}(x)}{\overline{F}_X(x)} = \infty$.*
2. *If $\sum_{n=1}^{\infty} np_n < \infty$, then the finite convergence given by*

$$\liminf_{x\to\infty} \frac{\overline{F_{Z_N}}(x)}{\overline{F}_X(x)} \to \sum_{n=1}^{\infty} np_n < \infty.$$

The second case is obtained when one of the following conditions is satisfied:

1. *If $\sum_{n=1}^{\infty} p_n(1+\varepsilon)^n < \infty$ for some $\varepsilon > 0$.*
2. *The tail of the severity distribution is regularly varying $\overline{F}_X \in RV_{-\rho}$ and can be represented by $\overline{F}_X(x) = x^{-\rho} l(x)$ with index $\rho > 1$ where $l(x)$ is a slowly varying function satisfying $\lim_{x\to\infty} l(ax)/l(x) = 1$ for every finite a and*

$$\mathbb{E}[N^{\rho+\varepsilon}] = \sum_{n=1}^{\infty} n^{\rho+\varepsilon} p_n < \infty$$

   for an arbitrarily small $\varepsilon > 0$.
3. *The tail of the severity distribution is regularly varying $\overline{F}_X \in RV_{-\rho}$ with index $\rho \in (0, 1)$.*
4. *The tail of the severity distribution, $\overline{F}_X$, is a member of the sub-family of functions which are O-regularly varying and satisfy for all $y \in \mathbb{R}$ that*

$$\lim_{x\to\infty} \frac{\overline{F}_X(x+y)}{\overline{F}_X(x)} = 1. \tag{7.127}$$

In addition, it is assumed that the lower Matuszewska index $\beta(\overline{F}_X) < -1$ *and that*

$$\mathbb{E}[N^{-\beta(\overline{F}_X)+1+\varepsilon}] < \infty$$

for an arbitrarily small $\varepsilon > 0$. *This means that the severity distribution* $\overline{F}_X$ *is a member of the family of functions that are formed as a subset of the sub-exponential distribution tails that comprise the intersection of the long-tailed functions and the O-regularly varying tails,* $\overline{F}_X \in OR \cap \mathscr{L}$.

5. *The tail of the severity distribution,* $\overline{F}_X$, *is a member of the sub-family of functions which are O-regularly varying and satisfy for all* $y \in \mathbb{R}$ *that*

$$\lim_{x\to\infty} \frac{\overline{F}_X(x+y)}{\overline{F}_X(x)} = 1, \tag{7.128}$$

and in addition, it is assumed that the lower Matuszewska index $\beta(\overline{F}_X) > -1$.

Furthermore, Daley *et al.* (2007) proved that when one considers an LDA model with compound Poisson-α-stable (where the α-stable severity model has support $[0, \infty)$], which occurs for the $S_\alpha(\beta, \gamma, \delta; S(0))$ parametrization when $\beta = 1$ and $\delta = 0$), then the ratio $\overline{F}_{Z_N}(x)/\overline{F}(x)$ is monotonically increasing in x for fixed n. Therefore, instead of the dominated convergence theorem, one may apply the monotone convergence theorem with the implications that one only requires as a consequence that $\mathbb{E}[N]$ is finite.

7.4.3 REMAINDER ANALYSIS: FIRST-ORDER SINGLE RISK LOSS PROCESS ASYMPTOTICS FOR HEAVY-TAILED LDA MODELS

In this section, the behaviour of the remainder term in the asymptotic first-order tail approximation of the LDA model is considered. As discussed in Omey and Willekens (1986), whilst the first-order dynamics have had numerous studies undertaken, there have been far fewer discussions about the second-order behaviour and in particular regarding the remainder term in the first-order asymptotic given by

$$R_1(x) := \overline{F}_{Z_N}(x) - \mathbb{E}[N]\overline{F}_X(x), \tag{7.129}$$

where the subscript R_1 indicates that the error is studied between the true annual loss distribution and the first-order approximation.

In Omey and Willekens (1986), two sets of results were obtained on the behaviour of the remainder term depending on whether the mean of the severity distribution was finite or not in the case of an assumption of regular variation of the tails of the severity distribution. In the case of a finite mean, the asymptotic behaviour of the remainder term can be linked to the behaviour of the derivative of the severity distribution. In the case of the severity distribution having heavy tails with an infinite mean, a different result arises as discussed later. The remainder term was also considered in the same way as in the work of Degen (2010, proposition A3) for a class of sub-exponential distributions, which will be considered first below.

The following refined asymptotic tail approximation for a sub-exponential single risk severity model is obtained in Theorem 8.3.

Proposition 7.15 (Remainder Analysis: Sub-exponential Severity Distribution) *Consider a severity distribution for losses* $X \sim f_X$ *which is zero at the origin* ($x = 0$) *and satisfies that both*

the tail distribution $\overline{F}_X$ and the density f_X are sub-exponential, $\overline{F}_X \in \mathcal{F}$ and $f_X \in \mathcal{F}$. Furthermore, assume that the frequency distribution is $N \sim F_N(n)$ such that its probability-generating function, given by

$$p_N(v) = \mathbb{E}[v^N] = \sum_{n=0}^{\infty} \Pr[N = n] v^n, \tag{7.130}$$

is analytic at $v = 1$. Then one has, for finite mean severity models ($\mathbb{E}[X] < \infty$), the remainder term $R_1(x) := \overline{F}_Z(x) - \mathbb{E}[N]\overline{F}_X(x)$ relative to the severity density satisfies asymptotically

$$\lim_{x \to \infty} \frac{R_1(x)}{f_X(x)} = \mathbb{E}[X]\mathbb{E}[(N-1)N]. \tag{7.131}$$

If the mean of the severity distribution is infinite and the density satisfies $f_X \in RV_{-1/\beta-1}$ for $1 \leq \beta < \infty$, then one has

$$\lim_{x \to \infty} \frac{R_1(x)}{f_X(x) \int_0^x \overline{F}_X(s)\, ds} = c_\beta \mathbb{E}[(N-1)N] \tag{7.132}$$

with $c_1 = 1$ and $c_\beta = (1-\beta)\frac{\Gamma^2(1-1/\beta)}{2\Gamma(1-2/\beta)}$ for $\beta \in (1, \infty)$.

Remark 7.8 *The second component of the above result for infinite mean severity models which are in the class of regular variation relates directly to the case studied in Omey and Willekens (1986).*

One of the key results studied for the case of i.i.d. losses $X_i \sim F_X$ with an infinite mean $\mathbb{E}[X_i]$ is to study the asymptotic behaviour of the total integrated error for the remainder given by

$$\lim_{x \to \infty} \int_0^x R(y)\, dy = \lim_{x \to \infty} \int_0^x (\overline{F}_{Z_N}(y) - \mathbb{E}[N]\overline{F}_X(y))\, dy. \tag{7.133}$$

Omey and Willekens (1986) showed the following result given in Theorem 7.10 for the case in which the severity distribution is assumed to have tails which are regularly varying. They also presented an analogous result in settings in which the severity model also admits a density function which is regularly varying.

Theorem 7.10 (Remainder Analysis: Regular Varying Severity Distribution) *If one considers an OpRisk LDA loss model with a severity distribution that satisfies that the tail $\overline{F}_X \in RV_{-\beta}$ for some $\beta \in [0, 1]$ and the frequency distribution satisfies that its probability-generating function*

$$\mathbb{E}[z^N] = \sum_{n=0}^{\infty} \Pr[N = n] z^n, \tag{7.134}$$

is analytic at $x = 1$ (not heavy tailed), then the remainder term of the annual loss distribution after removing the first-order approximation given by $R(x) = \overline{F}_Z(x) - \mathbb{E}[N]\overline{F}_X(x)$ will satisfy the result

$$\lim_{z \to \infty} \frac{\int_0^z R(y)\, dy}{\overline{F}_X(z) \int_0^z \overline{F}_X(y)\, dy} = \begin{cases} -2\mathbb{E}\left[\frac{N!}{2(N-2)!}\right], & \text{if } \beta = 1, \\ -(1-\beta)B(1-\beta, 1-\beta)\mathbb{E}\left[\frac{N!}{2(N-2)!}\right], & \text{if } \beta < 1, \end{cases} \tag{7.135}$$

with beta function $B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$. *Furthermore, if the severity density* $f_X(x)$ *exists and is regularly varying with* $\beta = -1$, *then one has*

$$R_1(x) \sim -\mathbb{E}\left[\frac{N!}{2(N-2)!}\right]\overline{F}_X(x)^2, \quad as \quad x \to \infty. \tag{7.136}$$

Remark 7.9 *One can show that if the severity distribution satisfies*

$$\mu = \int_0^\infty \overline{F}_X(y)dy < \infty, \tag{7.137}$$

then in the above result, one must have $\beta = 1$ *and the integrated total error in the first-order approximation is zero, that is,* $\int_0^\infty R_1(y)dy = 0$. *As noted in Omey and Willekens (1986), in this case, one would obtain from the above result*

$$\lim_{x\to\infty} \frac{\int_x^\infty R_1(y)dy}{1 - f_X(x)} = 2\mu\mathbb{E}\left[\frac{N!}{2(N-2)!}\right]. \tag{7.138}$$

Theorem 7.11 (Remainder Analysis for First Order Approximations: RV Density Case) *If one considers an OpRisk LDA loss model with a severity distribution that satisfies that the density exists and is regularly varying* $f_X \in RV_{-\alpha}$ *for some* $\alpha \in [1,2]$ *and the frequency distribution satisfies that its probability-generating function*

$$\mathbb{E}[z^N] = \sum_{n=0}^\infty \mathbb{P}\mathrm{r}[N=n]z^n \tag{7.139}$$

is analytic at $x = 1$ *(not heavy tailed), then the remainder term of the annual loss distribution after removing the first-order approximation given by* $R_1(x) = \overline{F}_Z(x) - \mathbb{E}[N]\overline{F}_X(x)$ *will satisfy one of the following results depending on the index of regular variation:*

$$\begin{aligned}
\lim_{x\to\infty} \frac{R_1(x)}{f_X(x)\int_0^x \overline{F}_X(y)dy} &= \begin{cases} c(\alpha)\mathbb{E}\left[\frac{N!}{2(N-2)!}\right], & if \quad \alpha \in (1,2), \\ 2\mathbb{E}\left[\frac{N!}{2(N-2)!}\right], & if \quad \alpha = 2, \end{cases} \\
\lim_{x\to\infty} \frac{R_1(x)}{\overline{F}_X(x)^2} &= -\mathbb{E}\left[\frac{N!}{2(N-2)!}\right], \quad if \quad \alpha = 1,
\end{aligned} \tag{7.140}$$

with function $c(\alpha)$ *given by*

$$c(\alpha) = \frac{(2-\alpha)\Gamma(2-\alpha)^2}{(\alpha-1)\Gamma(3-2\alpha)}. \tag{7.141}$$

Examples of results for such remainder terms in the α-stable severity model family is provided in Omey and Willekens (1986, lemma 2.5).

7.4.4 SUMMARY: FIRST-ORDER SINGLE RISK LOSS PROCESS ASYMPTOTICS FOR HEAVY-TAILED LDA MODELS

Consider the compound process distribution and right tail distribution functions, given for annual loss $Z_N = \sum_{i=1}^{N} X_i$ by

$$F_{Z_N}(x) = \sum_{n=0}^{\infty} p_n F_X^{(n)*}(x),$$

$$\overline{F}_{Z_N}(x) = \sum_{n=1}^{\infty} p_n \overline{F_X^{(n)*}}(x),$$

where it is assumed that the frequency probability mass function satisfies $\sum_{n=0}^{\infty} np_n < \infty$. Then the following first-order single risk loss processes tail asymptotic results apply when one considers a severity model with a regularly varying right tail.

Theorem 7.12 (First-Order Single Risk Asymptotic Tail Approximation) *The following results for compound process tail behaviour hold.*

1. *Assume that the distribution of X_i losses is in the sub-exponential severity family and an independent integer-valued random variable for the number of losses N satisfies $\mathbb{E}[(1+\varepsilon)^N] < \infty$, for an $\varepsilon > 0$. Then the following asymptotic equivalence holds:*

$$\Pr[Z_N > x] \sim \mathbb{E}[N]\Pr[X_1 > x]. \tag{7.142}$$

2. *Assume that the distribution of X_i is regularly varying with positive index $\rho > 0$ and the mean number of losses is finite $\mathbb{E}[N] < \infty$ and $\Pr[N > x] = o(\Pr[X_1 > x])$. Then the following asymptotic equivalence holds:*

$$\Pr[Z_N > x] \sim \mathbb{E}[N]\Pr[X_1 > x]. \tag{7.143}$$

3. *Assumes that the counting random variable N is regularly varying with index $\beta \geq 0$. If $\beta = 1$, then assume that $\mathbb{E}[N] < \infty$. Let the i.i.d. loss random variables (X_i) with finite mean $\mathbb{E}[X_i] < \infty$ and tail probability with asymptotic condition $\Pr[X_i > x] = o(\Pr[N > x])$, then asymptotically one has the equivalence for the compound processes*

$$\Pr[Z_N > x] \sim (\mathbb{E}[X_1])^{\beta}\Pr[N > x]. \tag{7.144}$$

4. *Assume that $\Pr[N > x] \sim c\Pr[X_1 > x]$ for some $c > 0$ and X_1 is regularly varying with an index $\rho \geq 1$ and with finite mean $\mathbb{E}[X_1] < \infty$. Then one has the asymptotic equivalence*

$$\Pr[Z_N > x] \sim \Pr[X_1 > x](\mathbb{E}[N] + c(\mathbb{E}[X_1])^{\rho}). \tag{7.145}$$

Remark 7.10 *The above results make intuitive sense, because, for example, the heavy-tailed sub-exponential severity model case, the tail of the distribution of the sum and the tail of the maximum are asymptotically of the same order. Hence, the sum of losses is governed by large individual losses, and, therefore, the distributional tail of the sum is mainly determined by the tail of the maximum loss. In addition, in the case of the general sub-exponential models, we note that the results presented for the first-order single-loss tail approximation required that $\mathbb{E}[z^N] < \infty$ for some $z > 1$. This means that N must have finite moments of all orders. This clearly has implications on the properties of the severity distribution because we observed the result that*

$$\lim_{x\to\infty}\frac{\overline{F}_{Z_N}(x)}{\overline{F}_X(x)}=\mathbb{E}[N]. \tag{7.146}$$

Remark 7.11 *The representation provided by the SLAs are remarkable as they show that for the annual loss distribution under an OpRisk LDA model for general sub-exponential family severity models, when considered at high confidence levels, the resulting quantiles of the annual loss distribution become only dependent on the tail of the severity distribution and not on the body. Therefore, when making such an asymptotic approximation, it is convenient that one only requires a quantification of the mean of the frequency distribution. Consequently, overdispersion as captured by negative binomial processes will not affect the high confidence level quantiles of the annual loss distribution.*

7.5 Refinements and Second-Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses

In this section, we refine the solutions derived above for first-order asymptotics by incorporating an additional term in the expansion under different tail assumptions for the severity models in the LDA model. This allows one to obtain a more refined and more accurate assessment of the annual loss tail behaviour at the expense of a greater complexity in the representation. The second-order SLA as discussed in Albrecher *et al.* (2010) and derived in Omey and Willekens (1986) takes the form given, for example, by Definitions 7.6 and 7.16.

In the following, three different second-order results are presented:

Case 1. The finite average loss and continuous long-tailed severity density setting;

Case 2. The regularly varying severity distribution and severity density setting; and

Case 3. The discontinuous severity density setting with bounded total variation.

Definition 7.6 (Second-Order Single-Loss Approximation: (Case 1)) *Given a sequence of i.i.d. losses X_i with sub-exponential severity distribution function f_X in which $\mathbb{E}[X]<\infty$. Furthermore, assuming that the severity distribution admits a continuous long-tailed severity density $f_X(x)$ which is dominantly varying with an upper Matuszewska index given by $\alpha(f_X)<-1$. Furthermore, assuming that there is also an independent integer-valued random variable for the number of losses, N, satisfying $\mathbb{E}[(1+\varepsilon)^N]<\infty$, for an $\varepsilon>0$. Then the second-order asymptotic approximation of the compound process tail probabilities, as $x\to\infty$, is given by*

$$\begin{aligned}\mathbb{P}\mathrm{r}[X_1+X_2+\cdots+X_N\geq x]&=\overline{F}_Z(x)\\&=a_2(x)+o(f_X(x)),\end{aligned} \tag{7.147}$$

with

$$\begin{aligned}a_2(x)&=a_1(x)+2\mathbb{E}\left[\frac{N!}{(N-2)!2!}\right]\mathbb{E}[X]f_X(x)\\&=\mathbb{E}[N]\overline{F}_X(x)+2\mathbb{E}\left[\frac{N!}{(N-2)!2!}\right]\mathbb{E}[X]f_X(x).\end{aligned} \tag{7.148}$$

If some additional restrictions are placed on the severity distribution function or the density, then the results presented in Proposition 7.16 owing to Omey and Willekens (1986) can be developed.

Proposition 7.16 (Second-Order Single-Loss Approximation (Case 2)) *Given a severity distribution function f_X which has regularly varying tail with index $\rho \in (0,1]$ such that $\overline{F}_X(x) \in RV_{-\rho}$ and a severity density $f_X(x)$ which is regularly varying, then the second-order approximation for $\overline{F}_Z(x)$ is given by*

$$a_2(x) = \begin{cases} a_1(x) - \dfrac{(2-\rho)\Gamma(2-\rho)}{(\rho-1)\Gamma(3-2\rho)}\mathbb{E}\left[\dfrac{N!}{(N-2)!2!}\right] f_X(x)\displaystyle\int_0^x \overline{F}_X(y)\,dy, & 0 < \rho < 1, \\ a_1(x) + 2\mathbb{E}\left[\dfrac{N!}{(N-2)!2!}\right] f_X(x)\int_0^x \overline{F}_X(y)\,dy, & \rho = 1. \end{cases}$$

If we define the total variation of a severity density f_X over an interval $[a,b]$ according to the definition provided in Equation 7.149, then one can state an alternative second-order approximation in Proposition 7.17 owing to Grübel (1987). This particular result is interesting for practitioners when models such as a splice severity model are adopted for modeling the main body of the severity losses and the tails separately. In such cases, it is common to admit discontinuities in the severity model.

Proposition 7.17 (Second-Order Single-Loss Approximation (Case 3)) *Consider the severity density $f_X(x)$ which is not continuous, but satisfies that the total variation is bounded on an interval $[x, x+1] \subset \mathbb{R}$, such that*

$$V_x^{x+1}(f) = \sup_P \sum_{i=0}^{n_P - 1} |f(x_{i+1}) - f(x_i)|, \tag{7.149}$$

where the supremum runs over the set of all partitions

$$\mathscr{P} = \{P = \{x_0, \ldots, x_{n_P}\} | P \text{ is a partition of } [a,b]\}$$

of the given interval. Then one can show that if there exists a monotonically decreasing function $\tau(x)$ with the property

$$\sup_{x>0} \frac{\tau(x)}{\tau(2x)} < \infty,\ \tau(x) = O(x^{-4}),\ \text{and } V_x^{x+1}(f) = o(\tau(x)), \tag{7.150}$$

then the second-order approximation can be written according to

$$\begin{aligned} \Pr[X_1 + X_2 + \cdots + X_N \geq x] &= \overline{F}_Z(x) \\ &= a_2(x) + o(\tau(x)), \end{aligned} \tag{7.151}$$

where $a_2(x)$ takes one of the forms above, depending on the tail assumptions made.

Hence, one can utilize these second-order approximations to make refinements to the first-order asymptotic expansion which are obtained under additional assumptions, see discussions in Albrecher *et al.* (2010).

Proposition 7.18 (Refined Second-Order Single-Loss Approximations) *Consider a severity distribution for losses $X \sim F_X$ which is zero at the origin ($x = 0$) and satisfies that both the tail distribution $\overline{F}_X$ and the density f_X are sub-exponential. Furthermore, assume the frequency distribution $N \sim F_N(n)$ is such that its probability-generating function given by*

$$p_N(v) = \mathbb{E}[v^N] = \sum_{n=0}^{\infty} \Pr[N = n] v^n \tag{7.152}$$

is analytic at $v = 1$, then one has, for finite mean severity models ($\mathbb{E}[X] < \infty$), that

$$\lim_{x \to \infty} \frac{\overline{F}_Z(x) - \mathbb{E}[N]\overline{F}_X(x)}{f_X(x)} = \mathbb{E}[X]\mathbb{E}[(N-1)N]. \tag{7.153}$$

If the mean of the severity distribution is infinite and the density satisfies $f_X \in RV_{-1/\beta-1}$ for $1 \leq \beta < \infty$, then one has

$$\lim_{x \to \infty} \frac{\overline{F}_Z(x) - \mathbb{E}[N]\overline{F}_X(x)}{f_X(x) \int_0^x \overline{F}_X(s)\, ds} = c_\beta \mathbb{E}[(N-1)N] \tag{7.154}$$

with $c_1 = 1$ and $c_\beta = (1-\beta)\frac{\Gamma^2(1-1/\beta)}{2\Gamma(1-2/\beta)}$ for $\beta \in (1, \infty)$.

Remark 7.12 *Note that many distributions will satisfy the conditions in Proposition 7.18 such as LogNormal, Weibull, Benktander Type I and Type II and numerous others discussed in Cruz* et al. *(2015, chapter 9) and Chapter 4.*

This section is concluded with some examples of first- and second-order tail asymptotic approximations in some LDA models that are closed form, as studied in Chapter 6, and, therefore, provide a benchmark framework to assess the asymptotic accuracy.

EXAMPLE 7.1 Single Risk LDA Poisson-Inverse-Gaussian Family

Consider the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N_t}$, and chosen to be an inverse Gaussian distribution $X_i \sim IGauss(\widetilde{\mu}, \widetilde{\lambda})$ where the two parameters in this model correspond to parameterizing the mean of the distribution for the severity $\widetilde{\mu}$ and the shape of the distribution $\widetilde{\lambda}$. Also, assume that the frequency is modeled as $N_t \sim Poisson(\lambda)$. The survival and quantile functions of the severity are given by

$$f(x; \widetilde{\mu}, \widetilde{\lambda}) = \left(\frac{\widetilde{\lambda}}{2\pi}\right)^{1/2} x^{-3/2} \exp\left\{-\frac{\widetilde{\lambda}(x-\widetilde{\mu})^2}{2\widetilde{\mu}^2 x}\right\}$$

$$F(x; \widetilde{\mu}, \widetilde{\lambda}) = \Phi\left(\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}} - 1\right)\right) + \exp\left(\frac{2\widetilde{\lambda}}{\widetilde{\mu}}\right)\Phi\left(-\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}} + 1\right)\right).$$

As the inverse Gaussian family is closed under convolution, that is,

$$X_i \sim IGauss(\widetilde{\mu}, \widetilde{\lambda}) \Rightarrow S_n = \sum_{i=1}^{n} X_i \sim IGauss(n\widetilde{\mu}, n^2\widetilde{\lambda}), \tag{7.155}$$

we can calculate the distribution of the compound process analytically (see the comparison with the approximations in Figure 7.1). The drawback of this family is that there is no closed form for the inverse cumulative distribution function (CDF), which obliges us to resort to a numerical procedure for obtaining the quantiles; fortunately, this is efficient and accurate for this class of models. For this model, we also present first- and second-order approximations for the value-at-risk (VaR) and spectral risk measure (SRM) for different choices of parameters in Figures 8.4 and 8.5. For the Poisson-inverse-Gaussian case, the greatest sensitivity is clearly on the frequency rate λ (see, e.g. the plot in Figure 8.4c). In difference to the LogNormal example, the second-order approximation seems to perform always better than the first order, but none of them are sufficiently close to any of the 'true' (Monte Carlo) VaR. These results, therefore, serve to illustrate the importance of understanding and developing further studies on the rate of convergence of these asymptotic SLAs. This will help to guide in the understanding of when they can be reliably utilized in practice.

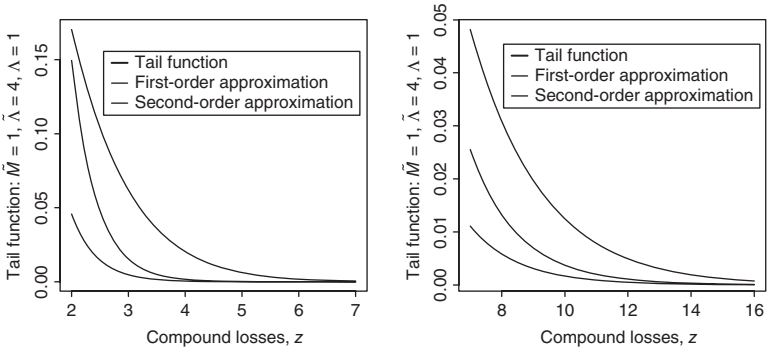

FIGURE 7.1 Tail function approximation for the Poisson-inverse-Gaussian example.

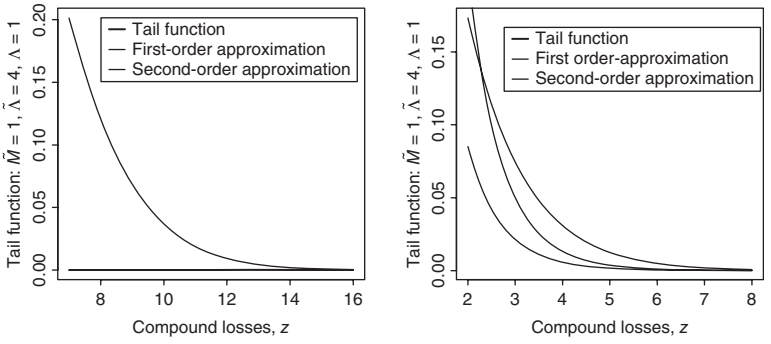

FIGURE 7.2 Tail function approximation for the Poisson-inverse-Gaussian example.

Next we consider the first- and second-order asymptotic tail approximations of the annual loss distributions that may be obtained under different assumptions regarding the dependence between the severities in the LDA model and their marginal tail behaviour.

7.6 Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Dependent Losses

Recently, there has been significant interest in understanding and studying the influence played by dependence between losses in the OpRisk LDA model and in particular its effect on the tail asymptotics. The focus of this section is to consider the combinations of joint dependence structure and marginal severity model tail features that will result in unchanged tail asymptotics for the LDA compound process model and partial sums, and, therefore, first- and second-order results applied earlier can be utilized. As is the case for the independent tail asymptotics, the majority of cases studied in the literature consider the marginal distributions of the loss random variables to have sub-exponential tail behaviour for the distribution, the exception being one of the earlier recent studies of this kind in which CV tail behaviour was considered, see Chen and Zhang (2007) and the precise large deviation results of Tang (2006).

Before presenting the asymptotic results, it is important to recall some basic properties of dependence that will be considered in defining tail asymptotics for dependent LDA model structures. To understand the role played by dependence, one must first develop appropriate methods to specify and model such dependence features of interest in LDA models. In this context, we will be primarily concerned with dependence between the losses in the compound process, and typically, we will treat the number of losses in the LDA model as independent from the severity model. Therefore, the severities in a year will come from a multivariate distribution with some form of dependence present. Quantifying this dependence in such a manner that each loss is no longer i.i.d. whilst ensuring the first- or second-order asymptotics are unaffected is a delicate issue and will depend on the form of dependence considered. In Cruz *et al.* (2015, chapter 10), a general discussion on possible dependence models is provided; in this section, we mention a few other notions of dependence based on ideas in Khursheed and Lai Saxena (1981) for quantification of positive dependence and in Block *et al.* (1982) for specifications of negative dependence. In general, there are also many ways of parameterizing and understanding distributional dependence through notions of copula, see Cruz *et al.* (2015, chapter 10, 11 and 12) and Marshall and Olkin (1988) and books such as Nelsen (1999) and Joe (1997).

In understanding the types of dependence which will not result in impact on the first- and second-order asymptotics, the most popular specifications proposed have been discussed later before presenting the results. This covers the following:

- Multivariate upper negative dependence (UND), lower negative dependence (LND) and negative dependence (ND);
- Multivariate association;
- Multivariate negative and positive quadrant dependence;
- Commonotonicity and stochastic ordering;
- Negative regression dependence;
- Parametric copula dependence.

We start with the concept of negative dependence first proposed in Block *et al.* (1982) and then studied further in Ghosh (1981), with the specification taken as given in Definition 7.7.

Definition 7.7 (Multivariate Negative Dependence in LDA Single Risk Models) *Consider a sequence of loss random variables in an OpRisk loss model* $\{X_i\}_{i\geq 1}$. *The sequence can be called LND or UND or negatively dependent as follows:*

- LND. *A sequence of loss random variables is LND if for each* $n \geq 1$ *and all* $X_1, X_2, \ldots, X_n$ *one has*

$$\mathbb{P}\mathrm{r}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n] \leq \prod_{i=1}^{n} \mathbb{P}\mathrm{r}[X_i < x_i]. \tag{7.156}$$

- UND. *A sequence of loss random variables is UND if for each* $n \geq 1$ *and all* $X_1, X_2, \ldots, X_n$ *one has*

$$\mathbb{P}\mathrm{r}[X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n] \leq \prod_{i=1}^{n} \mathbb{P}\mathrm{r}[X_i > x_i]. \tag{7.157}$$

- ND. *A sequence of loss random variables is negatively dependent if for each* $n \geq 1$ *and all* $X_1, X_2, \ldots, X_n$ *they satisfy that they are both LND and UND.*

Remark 7.13 (LND and UND versus Negative Association) *The notion of LND and UND is a weaker notion of dependence than the more familiar idea of negative association (NA).*

To make precise this remark, we recall in Definition 7.8 the concept of negative association of random variables, see, for instance, Joag-Dev and Proschan (1983).

Definition 7.8 (Multivariate Negative Association in LDA Single Risk Models) *Consider a sequence of loss random variables in an OpRisk loss model* $\{X_i\}_{i\geq 1}$. *The sequence is called negatively associated (NA) if for every pair of disjoint subsets* A_1, A_2 *of* $\{1, 2, \ldots, n\}$ *one has*

$$\mathbb{C}\mathrm{ov}[f_1(X_i, i \in A_1), f_2(X_j, j \in A_2)] \leq 0 \tag{7.158}$$

whenever f_1 *and* f_2 *increasing functions.*

Remark 7.14 *Practically important multivariate distributions with the property of negatively association include multinomial, multivariate hypergeometric, Dirichlet and Dirichlet compound multinomial distributions.*

The following properties of negatively associated (NA) random sequences of loss random variables given in Proposition 7.19 is also useful to recall, see Joag-Dev and Proschan (1983).

Proposition 7.19 (Properties of Negatively Associated Loss Random Variables) *Consider a sequence of NA loss random variables in an OpRisk loss model* $\{X_i\}_{i\geq 1}$, *then the following properties apply:*

- *A subset of two or more NA random variable losses is NA.*

- *A set of independent random variable losses is NA.*
- *Increasing functions defined on disjoint subsets of a set of NA random variable losses are NA.*
- *Unions of independent sets of NA random variable losses are NA.*

A third notion of dependence that is of significance in the following sets of results relates to the notion of pairwise quadrant dependence given by Definition 7.9, see Lehmann (1966) and the associated pairwise positive quadrant dependence (PPQD).

Definition 7.9 (Pairwise Negative Quadrant Dependence) *A pair of loss random variables X_i and X_j are said to be pairwise negative quadrant dependent (PNQD) if for all $x, y \in \mathbb{R}$ one has*

$$\mathbb{P}\mathrm{r}[X_i \leq x, X_j \leq y] \leq \mathbb{P}\mathrm{r}[X_i \leq x]\mathbb{P}\mathrm{r}[X_j \leq y]. \tag{7.159}$$

Definition 7.10 (Pairwise Positive Quadrant Dependence) *A pair of loss random variables X_i and X_j are said to be PPQD if for all $x, y \in \mathbb{R}$ one has*

$$\mathbb{P}\mathrm{r}[X_i \leq x, X_j \leq y] \geq \mathbb{P}\mathrm{r}[X_i \leq x]\mathbb{P}\mathrm{r}[X_j \leq y]. \tag{7.160}$$

The fourth approach that has been adopted in considering dependence structures and their influence on compound process tail asymptotics is due to Ko and Tang (2008, assumption 2.1) where they considered the bivariate setting and the general assumption on dependence given in Definition 7.11. This dependence assumption is based on setting up a conditional probability in a ratio, such that if one of the variables were independent, then the ratio should collapse to 1. This general dependence representation allows one to capture limited positive dependence as well as ND features in particular limited positive and negative quadrant dependence.

Definition 7.11 (Negative Regression Dependence: Bivariate) *Consider two loss random variables X_1 and X_2 in which X_2 is stochastically decreasing in X_1 such that $SD(X_2|X_1)$ and define their dependence relationship for $(i, j) \in \{(1, 2), (2, 1)\}$ by*

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_j > x - y | X_i = y]}{\mathbb{P}\mathrm{r}[X_j > x - y]} = O(1) \tag{7.161}$$

which will hold uniformly for all $y \in [x_0, x]$ for some large $x_0 > 0$ such that

$$\limsup_{x\to\infty} \sum_{x_0 \leq y \leq x - x_0} \frac{\mathbb{P}\mathrm{r}[X_j > x - y | X_i = y]}{\mathbb{P}\mathrm{r}[X_j > x - y]} < \infty. \tag{7.162}$$

The following relationship will hold, for all $x_0 \geq 0$ such that $F_{X_1}(x_0) > 0$, uniformly for all $y > x_0$ according to

$$\mathbb{P}\mathrm{r}[X_2 > x - y | X_1 = y] \leq \frac{\overline{F}_{X_2}(x - y)}{F_{X_1}(x_0)}. \tag{7.163}$$

Before presenting the multivariate analogue of the notion of negative regression dependence, it is first instructive to recall the notion of comonotonicity and stochastic order of first order.

Remark 7.15 (Comonotonicity and Stochastically Decreasing) *The dependence relation imposed by Definition 7.11 excludes any extremely positive dependence structures such as those arising from comonotonic (non-decreasing support) loss random variables which satisfy*

$$\mathbb{P}\mathrm{r}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n] = \min_{i \in \{1,2,\ldots,n\}} F_{X_i}(x_i). \tag{7.164}$$

Definition 7.12 (Negative Regression Dependence: Multivariate) *Consider n loss random variables with marginal distributions $X_i \sim F_{X_i}$ and a joint dependence that is the analogue of the bivariate negative regression dependence for $n \geq 2$ that is captured by the following relationship on the conditional distributions of the partial sums:*

$$\frac{\mathbb{P}\mathrm{r}[Z_{j-1} > x - y | X_j = y]}{\mathbb{P}\mathrm{r}[Z_{j-1} > x - y]} = O(1), \tag{7.165}$$

which will hold uniformly for all $y \in [x_0, x]$ which should exist for some large $x_0 > 0$ such that this order of asymptotic convergence is satisfied for all $j \in \{2, \ldots, n\}$.

Note that Geluk and Tang (2009) considered a related approach to specification of the dependence in the severity loss random variables under two assumptions (A1) and (A2). They showed that such dependence relationships, under their version of the proposed dependence in Definition 7.12, can be satisfied by a family of copula models such as the Farlie-Gumbel-Morgenstern (FGM) distribution, which is characterized in the following section after recalling the notion of a copula model. The two dependence specifications they developed involved for n-losses:

Assumption (A1) for all $i, j \in \{1, 2, \ldots, n\}$ one has the limit

$$\lim_{x_i \wedge x_j \to \infty} \mathbb{P}\mathrm{r}[|X_i| > x_i | X_j > x_j] = 0.$$

Assumption (A2) there exists x_0 and $c > 0$ such that one has the bound

$$\mathbb{P}\mathrm{r}[|X_i| > x_i | X_j = x_j] < c\overline{F}_{X_i}(x_i)$$

for all $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$ and $x_j > x_0$.

The fifth approach involves the specification of a particular parametric form of dependence via, for instance, a copula model specification, see Cruz *et al.* (2015, chapters 10, 11 and 12) for an in-depth discussion of copula model details. Here we provide a few additional results that will be of relevance when considering the copula modeling for asymptotics of dependent compound processes.

The definition of an n-dimensional copula distribution, denoted generically by $C(u_1, u_2, \ldots, u_n)$, was given as any distribution taking support on the unit d-hypercube that satisfies the following two conditions, see Roger (2006, definition 2.10.6):

1. For every vector $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in [0, 1]^n$, one can show that $C(\boldsymbol{u}) = 0$ if at least one co-ordinate of $\boldsymbol{u}$ is 0.
2. In addition, for every $\boldsymbol{a}$ and $\boldsymbol{b}$ in $[0, 1]^n$, such that $\boldsymbol{a} \leq \boldsymbol{b}$, i.e. $a_i < b_i$ for all $i \in \{1, 2, \ldots, n\}$, the following condition on the volume for copula C is satisfied, $V_C([\boldsymbol{a}, \boldsymbol{b}]) \geq 0$. In this notation, the volume of an n-box is given by

$$V_C([\boldsymbol{a}, \boldsymbol{b}]) = \sum \operatorname{sgn}(\boldsymbol{v}) C(\boldsymbol{v}) = \Delta_{a_1}^{b_1} \Delta_{a_2}^{b_2} \cdots \Delta_{a_n}^{b_n} C(\boldsymbol{v}), \tag{7.166}$$

where the sum is taken over all vertices $\boldsymbol{v}$ of the n-box $[\boldsymbol{a}, \boldsymbol{b}]$ and $\operatorname{sgn}(\boldsymbol{v}) = 1$ if $v_k = a_k$ for an even number of k's of $\operatorname{sgn}(\boldsymbol{v}) = -1$ if $v_k = a_k$ for an odd number of k's. In addition, one defines the notation

$$\Delta_{a_k}^{b_k} C(\boldsymbol{u}) = C(u_1, u_2, \ldots, u_{k-1}, b_k, u_{k+1}, \ldots, u_n) - C(u_1, u_2, \ldots, u_{k-1}, a_k, u_{k+1}, \ldots, u_n). \tag{7.167}$$

A very popular family of copula models in risk and insurance modeling is the Archimedean copula family, see discussion in Savu and Trede (2006) and McNeil and Nešlehová (2009). We recall their basic definition by generator functions below, see more detailed discussion in Cruz *et al.* (2015, chapters 10, 11 and 12). In particular, in the context of modeling dependence compound process with unaffected compound process first- and second-order tail asymptotics, a particular member of the Archimedean family will be of interest, the Ali-Mikhail-Haq (AMH) copula model.

Definition 7.13 (Archimedean Generator Function) *An Archimedean generator is a continuous, decreasing function* $\psi : [0, \infty] \to [0, 1]$ *which satisfies the following conditions:*

1. $\psi(0) = 1$;
2. $\psi(\infty) = \lim_{t \to \infty} \psi(t) = 0$; and
3. ψ *is strictly decreasing on* $[0, \inf\{t : \psi(t) = 0\}]$.

From the definition of the generator, one may define the family of Archimedean copula densities as follows.

Definition 7.14 (Archimedean Copula) *A* d*-dimensional copula* C *is called Archimedean if for some generator* ψ *it can be represented as*

$$C(\boldsymbol{u}) = \psi\{\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)\} = \psi\{t(\boldsymbol{u})\} \quad \forall \boldsymbol{u} \in [0, 1]^d, \tag{7.168}$$

where $\psi^{-1} : [0, 1] \to [0, \infty]$ *is the inverse generator with* $\psi^{-1}(0) = \infty\{t : \psi(t) = 0\}$ *and* $t(\boldsymbol{u}) = \psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)$. *The Archimedean copula* C *admits a density* c *if and only if* $\psi^{(d-1)}$ *exists and is absolutely continuous on* $(0, \infty)$. *When this condition is satisfied, the copula density* c *is given by*

$$c(\boldsymbol{u}) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d} = \psi^{(d)}\{t(\boldsymbol{u})\} \prod_{j=1}^{d} (\psi^{-1})'(u_j) \quad , \quad \boldsymbol{u} \in (0, 1)^d. \tag{7.169}$$

As discussed in Cruz *et al.* (2015, chapters 10, 11 and 12), there are many sub-families of Archimedean copula; here we just recall the definition of one of them that is important, the AMH copula given in Definition 7.15, see Kumar (2010).

Definition 7.15 (Ali-Mikhail-Haq Copula Models) *The AMH copula distribution defined in the bivariate case is given by*

$$C(u_1, u_2) = \frac{u_1 u_2}{1 - \rho(1 - u_1)(1 - u_2)} \tag{7.170}$$

for copula parameter $\rho \in [-1, 1]$. *In addition the AMH copula parameter* ρ *has the following relationship to Kendall's* τ *and Spearman's* ρ *rank correlations*

$$\begin{aligned} \tau &= \frac{3\rho - 2}{3\rho} - \frac{2(1-\rho)^2 \ln(1-\rho)}{3\rho^2}, \\ \rho &= \frac{12(1+\rho)\, di\{\ln\}(1-\rho) - 24(1-\rho)\ln(1-\rho)}{\rho^2} - \frac{3(\rho+12)}{\rho} \end{aligned} \tag{7.171}$$

with $di\{\ln\}(x) := \int_1^x \frac{\ln t}{1-t}\, dt$.

Having defined the n-copula, we also briefly recall the notion of tail dependence as it will play an important role in aspects of the study of the asymptotics of compound processes with copula dependence structures; there are several variants of tail dependence measures, see discussions in Cruz *et al.* (2015, chapters 10, 11 and 12) and Coles *et al.* (1999). In Definition 7.16, we provide some recent proposals for quantifying tail dependence under copula models, see discussions in De Luca and Rivieccio (2012), Coles *et al.* (1999) and Frahm *et al.* (2005).

Definition 7.16 (Tail dependence measures) *There are numerous ways to define tail dependence; the most popular involves considering two loss random variables* X_1 *and* X_2 *with marginal distributions* $F_{X_i}, i = 1, 2$ *and joint copula distribution* C*; one may define the coefficient of upper tail dependence by*

$$\lambda_u := \lim_{u \uparrow 1} \mathbb{P}\mathrm{r}\left[X_2 > F_{X_2}^{-1}(u) | X_1 > F_{X_1}^{-1}(u)\right] = \lim_{u \uparrow 1} \frac{1 - 2u + C(u, u)}{1 - u}, \tag{7.172}$$

and similarly, we define the coefficient of lower tail dependence by

$$\lambda_l := \lim_{u \downarrow 0} \mathbb{P}\mathrm{r}\left[X_2 \leq F_{X_2}^{-1}(u) | X_1 \leq F_{X_1}^{-1}(u)\right] = \lim_{u \downarrow 0} \frac{C(u, u)}{u}, \tag{7.173}$$

see discussions on these tail dependence definitions in Frahm et al. *(2005). In Coles* et al. *(1999), a modified version of upper tail dependence was considered according to*

$$\widetilde{\lambda}_u := \lim_{u \uparrow 1} \frac{2 \ln \mathbb{P}\mathrm{r}\left[U_1 > u\right]}{\ln \mathbb{P}\mathrm{r}\left[U_1 > u, U_2 > u\right]} - 1, \tag{7.174}$$

which when it exists is in the domain $\widetilde{\lambda}_u \leq 0$. *In De Luca and Rivieccio (2012), the following generalized multivariate expressions for upper tail dependence were obtained during their study of Archimedean copulae. Consider a vector of* d *loss random variables* $X = (X_1, ..., X_d)^T$ *with marginal loss distribution functions* $F_{X_1}, ..., F_{X_d}$ *and a joint dependence structure given by a copula*

C. The generalized coefficient of upper tail dependence is defined as follows (and illustrated for the Archimedean family):

$$\begin{aligned}
&\lambda_u^{1,\ldots,h|h+1,\ldots,d} \\
&= \lim_{\nu \to 1-} \Pr[X_1 > F^{-1}(\nu), \ldots, X_h > F^{-1}(\nu) | X_{h+1} > F^{-1}(\nu), \ldots, X_d > F^{-1}(\nu)] \\
&= \lim_{t \to 0^+} \frac{\sum_{i=1}^{d} \left(\binom{d}{d-i} i(-1)^i [\psi^{-1'}(it)] \right)}{\sum_{i=1}^{d-h} \left(\binom{d-h}{d-h-i} i(-1)^i [\psi^{-1'}(it)] \right)}.
\end{aligned} \tag{7.175}$$

The equivalent generalized coefficient for the lower tail dependence is defined as follows (and illustrated for the Archimedean family) according to

$$\begin{aligned}
&\lambda_l^{1,\ldots,h|h+1,\ldots,d} \\
&= \lim_{\nu \to 0+} \Pr[X_1 < F^{-1}(\nu), \ldots, X_h < F^{-1}(\nu) | X_{h+1} < F^{-1}(\nu), \ldots, X_d < F^{-1}(\nu)] \\
&= \lim_{t \to \infty} \frac{d}{d-h} \frac{\psi^{-1'}(dt)}{\psi^{-1'}((d-h)t)}.
\end{aligned} \tag{7.176}$$

The lower and upper tail dependence of the AMH copula family was given in Kumar (2010) according to the result in Proposition 7.20.

Proposition 7.20 (Upper and Lower Tail Dependence for AMH Copula) *Consider the bivariate copula distribution in the AMH family with parameter $\rho \in [-1, 1]$, then the upper and lower tail dependence are given by*

$$\begin{aligned}
\lambda_l &= \lim_{u \downarrow 0} \frac{C(u,u)}{u} = \begin{cases} 0.5, & \text{if } \rho = 1, \\ 0, & \text{if } \rho < 1, \end{cases} \\
\lambda_u &= \lim_{u \uparrow 1} \frac{1 - 2u + C(u,u)}{1-u} = 0.
\end{aligned} \tag{7.177}$$

This result shows that under an AMH copula, two loss random variables will be asymptotically dependent only if the copula parameter is on the boundary $\rho = 1$, otherwise they are asymptotically independent. Clearly, this shows that, in practice, one must be careful to undertake estimation of such parameters with care as we will see below that one value may admit an asymptotic compound process expansion, whilst others will not. The following properties of tail dependence are recalled for use in future sections of this chapter, see Albrecher *et al.* (2006).

Proposition 7.21 (Properties of Tail Dependence) *Consider two loss random variables with marginal loss distributions $X_i \sim F_{X_i}$ and a joint dependence modeled by the copula C, then defining the constant*

$$c = \lim_{x \to \infty} \frac{\overline{F}_{X_2}(x)}{\overline{F}_{X_1}(x)}, \tag{7.178}$$

one can show the following features of upper tail dependence:

1. *The following bounds apply to the upper tail dependence*

$$c\lambda_u \leq \widehat{\lambda} \leq \min(c, \lambda_u) \tag{7.179}$$

with

$$\widehat{\lambda} = \lim_{x\to\infty} \frac{1 - F_{X_1}(x) - F_{X_2}(x) + C(F_{X_1}(x), F_{X_2}(x))}{1 - F_{X_1}(x)}. \tag{7.180}$$

2. *The following relationship between the maximum of a sum of two random variables and the tail dependence holds*

$$\mathbb{P}\mathrm{r}[\max\{X_1, X_2\} > x] \sim (1 + c - \widehat{\lambda})\overline{F}_{X_1}(x) \tag{7.181}$$

and the tail result is given by

$$\lim_{x\to\infty} \mathbb{P}\mathrm{r}[X_1 > x | \max\{X_1, X_2\} > x] = \frac{1}{1 + c - \widehat{\lambda}}. \tag{7.182}$$

3. *The following worst case bounds can be obtained*

$$\overline{F}_{X_1}(x) << \mathbb{P}\mathrm{r}[X_1 + X_2 > x] << (1 + c)\overline{F}_{X_1}\left(\frac{x}{2}\right). \tag{7.183}$$

4. *Consider the identically distributed losses $X_i \sim f_X(x)$ with a copula distribution $C(u_1, u_2) = C(f_X(x), f_X(y))$, then one can obtain the following bounds:*

$$\lambda_u \leq \liminf_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[c_1X_1 + c_2X_2 > x]}{\mathbb{P}\mathrm{r}\left[X_1 > \frac{x}{c_1+c_2}\right]} \limsup_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[c_1X_1 + c_2X_2 > x]}{\mathbb{P}\mathrm{r}\left[X_1 > \frac{x}{c_1+c_2}\right]} \leq 2 - \lambda_u, \tag{7.184}$$

for constants c_1 and c_2 satisfying $y = \frac{c_1 x}{(c_1+c_2)}$.

One can also relat the notion of upper tail dependence to negative regression dependence as follows.

Remark 7.16 (Upper Tail Dependence and Negative Regression Dependence) *One can show that if loss random variables satisfy the definition negative regression dependence then they will always have upper tail dependence of zero as measured by the tail dependence measure*

$$\lambda_u = \lim_{u\uparrow 1} \mathbb{P}\mathrm{r}[U_2 > u | U_1 > u] \tag{7.185}$$

for marginally uniform $U_1 = F_{X_1}(x)$ and $U_2 = F_{X_2}(x)$.

Theorem 7.13 (Negative Regression Dependence via a Copula: Bivariate Case) *Consider two loss random variables X_1 and X_2 with marginal loss distributions $X_i \sim F_{X_i}$ and a copula dependence given by the copula distribution C. The joint distribution of the loss random variables $F_{X_1,X_2}(x_1, x_2)$ will satisfy that X_1 and X_2 have negative regression dependence if one rewrites the definition in one of the two ways in terms of the copula distribution:*

1. *If the first derivative of the copula distribution exists, then one can express the conditions on the copula to satisfy negative regression dependence by considering for* $u_1 = F_{X_1}(y)$ *and* $u_2 = F_{X_2}(x-y)$ *the relationship*

$$\frac{1 - \dfrac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2)}{\overline{F}_{X_2}(x-y)} = O(1). \tag{7.186}$$

This holds uniformly for all $y \in [x_0, x]$ *for some large* $x_0 > 0$.

2. *The alternative way of expressing the conditions on a copula distribution to satisfy negative regression dependence involves the copula density when it exists given by the mixed second-order partial derivatives:*

$$c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2). \tag{7.187}$$

If the copula density exists and it is uniformly bounded by a constant $M > 0$ *for all* $(u_1, u_2) \in [c, 1] \times [c, 1]$*, then the condition required for negative regression dependence is given by the relationship*

$$\frac{\int_{x-y}^{\infty} c(F_{X_1}(y), F_{X_2}(z))\, dF_{X_2}(z)}{\overline{F}_{X_2}(x-y)} \leq M. \tag{7.188}$$

In Geluk and Tang (2009), it was shown that their notion of negative regression dependence in the multivariate setting was satisfied by the family of Fralie-Gumbel-Morgenstern (FGM) copula models presented in Definition 7.17, see Johnson *et al.* (2002).

Definition 7.17 (Farlie-Gumbel-Morgenstern Copula Models) *Consider a loss process with* n *loss random variables with marginal distributions* $X_i \sim F_{X_i}(x)$ *and a joint distribution given by a copula* C*. The joint dependence will be in the family of FGM copulas with parameters* $\{a_{i,j}\}_{i,j=1}^{n}$ *which satisfy constraints provided in Johnson* et al. *(2002) and have a distribution given by*

$$F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \overline{F}_{X_i}(x_i) \left(1 + \sum_{1 \leq i < j \leq n} a_{i,j} \overline{F}_{X_i}(x_i) \overline{F}_{X_j}(x_j) \right).$$

One last copula we will utilize below is the Gaussian copula, so we briefly recall its specification in Definition 7.18 for the simple bivariate case, see most standard books on copula models for general extensions, for instance Joe *et al.* (2012).

Definition 7.18 (Gaussian Copula) *A bivariate copula is a Gaussian copula if it has a representation*

$$C(u, v) := \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)), \tag{7.189}$$

where

$$\Phi_\rho(x, y) := \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(\frac{2\rho st - s^2 - t^2}{2(1-\rho^2)} \right) dsdt$$

and Φ *denotes the standard Normal distribution function.*

7.6.1 SEVERITY DEPENDENCE STRUCTURES THAT DO NOT AFFECT LDA MODEL TAIL ASYMPTOTICS: STOCHASTIC BOUNDS

In this section, the main focus is to consider both partial sum and compound process asymptotics that will be unaffected by certain types of dependence. As was the case for the independent LDA model structures, the dependent cases can also be studied for a range of different heavy-tailed features. In the following section, a few of these results are highlighted when they relate to the first-order and second-order asymptotics presented previously.

In this section, the partial sums of $n \geq 1$ loss random variables are considered in which the losses are no longer i.i.d.

$$Z_n = \sum_{i=1}^{n} X_i \sim F_{Z_n}(x), \tag{7.190}$$

where the marginal distribution of each loss $X_i \sim F_{X_i}(x)$ could differ (although practically this is not typically relevant so one may restrict to the case of identical severity distribution) and each marginal satisfies that it was heavy tailed in some sense and the losses are dependent in some sense to be made precise. Note that in such settings, one can no longer obtain the joint distribution of the compound process simply as a mixture of the n-fold convolutions given by the distribution for each number of losses weighted by the probability of this many losses, because the standard convolution approach requires the i.i.d. properties of the summed random losses. In this section, the primary question considered involves addressing the following question. *To what extent and how can one incorporate dependence into the severity models of the LDA structure with the effect that the tail asymptotics are unaffected*? In particular, the asymptotic relationship given by

$$\begin{aligned} \overline{F}_{Z_n}(x) &= \Pr[X_1 + X_2 + \cdots + X_n > x] \\ &\sim \Pr[\max\{X_1, X_2, \ldots, X_n\} > x] \sim n\overline{F}_X(x) \end{aligned} \tag{7.191}$$

still holds, that is, one can still apply the results of Sgibnev (1996) which were based on the Pollaczek–Spitzer identity.

This question is very interesting because it was shown in Albrecher *et al.* (2006, theorem 2.10) that even in the simplest possible cases of bivariate partial sums of marginally heavy-tailed sub-exponential severity distributions, this is a non-trivial question to answer. In particular, Albrecher *et al.* (2006, theorem 2.10) states in the bivariate case that one may construct a multivariate distribution for the loss random variables such that the marginal loss distributions are heavy tailed and a copula distribution exists such that even for upper tail dependence between the severities $\lambda_u = 0$, one still has that the first-order asymptotic result for the partial sum does not hold, see Theorem 7.14. However, it will also be demonstrated in the following section that if the tails of the marginal distribution are regularly varying, then the condition of tail independence with $\lambda_u = 0$ suffices to ensure the first-order asymptotic result applies. Therefore, there is an interesting question to consider in how heavy should the tails be to ensure asymptotic validity of the asymptotic for partial sums.

Theorem 7.14 *Consider two losses X_1 and X_2 with marginal distribution given by F_X that satisfies the membership of the sub-family of sub-exponential distributions corresponding to those in the domain of attraction of the Gumbel minimum detectable activity (MDA), $F_X \in \mathcal{F} \cap MDA\,(Gumbel)$, then one can show that there exists a copula with upper tail dependence $\lambda_u = 0$ such that*

$$\lim_{x \to \infty} \frac{\Pr[X_1 + X_2 > x]}{\Pr[X_1 > x]} = \infty. \tag{7.192}$$

Remark 7.17 *This result shows that one may not automatically assume that with zero upper tail dependence, this will result in the first-order tail asymptotic for a partial sum to be valid.*

Before presenting the results for different types of dependence and their role on the compound process asymptotics, it will be instructive to first see the results for bounds on random sums obtained in Denuit *et al.* (1999) and then the bivariate asymptotic results studied in Albrecher *et al.* (2006). It can be shown that one may obtain stochastic bounds on the distribution of the partial sums Z_n for any $n \geq 1$ which are interesting in this section as they make no assumptions on the form of dependence that may be present between the loss random variables. These bounds rely on the concept of stochastic dominance as detailed in Definition 7.19, see Lehmann (1966). Note that stochastic dominance is a form of stochastic ordering.

Definition 7.19 (First-Order Stochastic Dominance) *Consider two loss random variables $X_1 \sim F_{X_1}(x)$ and $X_2 \sim F_{X_2}(x)$. One can say the X_1 first-order stochastically dominates X_2 if $F_{X_1}(x) \leq F_{X_2}(x)$ for all x.*

In the work of Denuit *et al.* (1999), the notion of stochastic dominance is utilized to obtain the following stochastic bounds on the partial sums given in Theorem 7.15.

Theorem 7.15 (Stochastic Bounds for Distributions of Partial Sums) *Consider the partial sum with n-losses, such that each loss is marginally given by a distribution $X_i \sim F_{X_i}$ with any form of desired dependence structure between the individual losses. Then the distribution for the partial sum of these loss random variables $Z_n = X_1 + X_2 + \cdots + X_n$ is stochastically bounded as follows:*

$$F_{\min}(z) \leq \Pr[Z_n \leq z] \leq F_{\max}(z), \quad z \in \mathbb{R}, \tag{7.193}$$

with the definitions:

$$\begin{aligned}
\Omega(z) &= \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n = z\}, \\
F_{\min}(z) &= \sup_{\boldsymbol{x} \in \Omega(z)} \max \left[\sum_{i=1}^{n} F_{X_i}(x_i) - (n-1), 0 \right], \\
F_{\max}(z) &= \inf_{\boldsymbol{x} \in \Omega(z)} \min \left[\sum_{i=1}^{n} F_{X_i}(x_i), 1 \right].
\end{aligned} \tag{7.194}$$

Note that some examples based on the generalized Pareto distribution (GPD) marginal distributions and discussion on computing these upper and lower bounds by optimization over the hyperplane are presented in Denuit *et al.* (1999) and Frank *et al.* (1987). In the bivariate case, the results provided in Theorem 7.16 relate to the refined bounds also provided in Denuit *et al.* (1999) for settings in which one considers tightening the bounds using additional assumptions or information on the structure of the dependence between the loss random variables.

Theorem 7.16 (Bounds on the Sum of Dependent Bivariate Losses) *Consider two marginally i.i.d loss random variables X_1 and X_2 such that the joint distribution $F_{X_1,X_2}(x_1, x_2)$ is bounded below by some other distribution function $G(x_1, x_2)$ for any $x_1, x_2 \geq 0$, then the following lower and upper bounds apply for the joint tail of the partial sum, for positive constants $c_1, c_2 > 0$:*

- *The lower bound of the bivariate sum is given by*

$$1-\inf_{y\geq 0}\left(F\left(\frac{y}{c_1}\right)+F\left(\frac{x-y}{c_2}\right)-G\left(\frac{y}{c_1},\frac{x-y}{c_2}\right)\right)\leq \Pr[c_1X_1+c_2X_2>x]. \tag{7.195}$$

- *The upper bound of the bivariate sum is given by*

$$\Pr[c_1X_1+c_2X_2>x]\leq 1-\sup_{y\geq 0}G\left(\frac{y}{c_1},\frac{x-y}{c_2}\right). \tag{7.196}$$

As shown in Denuit *et al.* (1999), for each given x value, these bounds are tight; however, it turns out to be more productive in the context of asymptotic analysis to consider specification explicitly of the form of dependence. The bivariate case was also extended in Denuit *et al.* (1999) to the n-variate partial sum as shown in Theorem 7.17.

Theorem 7.17 (Refined Partial Sum Stochastic Bounds) *Consider $X_1, X_2, \ldots, X_n$ loss random variables each with distribution $X_i \sim F_{X_i}$. If there exists a joint distribution $G(x_1, \ldots, x_n)$ that satisfies that if a lower bound on the joint distribution of the loss random variables*

$$G(x_1, \ldots, x_n) \leq \Pr[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n] \tag{7.197}$$

as well as a joint survival function $\overline{H}$ that satisfies

$$\Pr[X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n] \geq \overline{H}(x_1, x_2, \ldots, x_n) \tag{7.198}$$

for all $x_1, x_2, \ldots x_n \in \mathbb{R}$, then the stochastic bound is given by

$$\sup_{\boldsymbol{x}\in\Omega(z)} G(x_1, x_2, \ldots, x_n) \leq \Pr[Z_n \leq z] \leq 1 - \sup_{\boldsymbol{x}\in\Omega(z)} \overline{H}(x_1, x_2, \ldots, x_n) \tag{7.199}$$

for $z \in \mathbb{R}$ and $\Omega(z) = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 + x_2 + \cdots + x_n = z\}$.

Albrecher *et al.* (2006) demonstrated that if one considers the bivariate case, then one can obtain a lower bound for the tail probabilities of the partial sum in terms of the marginal tail probability and the dependence feature as quantified through a tail dependence measure, see Theorem 7.18

Theorem 7.18 (Bounding Tail Probabilities for Partial Sums of RV Severities) *Consider two loss random variables X_1 and X_2 which satisfy that the first has a distribution with tail property given by $\overline{F}_{X_1} \in RV_\alpha$ for $\alpha > 0$. Then one can show that the lower bound on the tail probability applies*

$$\limsup_{x\to\infty}\frac{\Pr[X_1+X_2>x]}{\overline{F}_{X_1}(x)}\leq\begin{cases}\left(\widehat{\lambda}^{\frac{1}{\alpha+1}}+\left(1+c-2\widehat{\lambda}\right)^{\frac{1}{\alpha+1}}\right)^{\alpha+1}, & \widehat{\lambda}\in\left[0,\frac{1+c}{3}\right],\\ 2^{\alpha}(1+c-\widehat{\lambda}), & \widehat{\lambda}\in\left[\frac{1+c}{3},1\right],\end{cases}$$

with

$$\widehat{\lambda}=\lim_{x\to\infty}\frac{1-F_{X_1}(x)-F_{X_2}(x)+C(F_{X_1}(x),F_{X_2}(x))}{1-F_{X_1}(x)}. \tag{7.200}$$

7.6.2 SEVERITY DEPENDENCE STRUCTURES THAT DO NOT AFFECT LDA MODEL TAIL ASYMPTOTICS: SUB-EXPONENTIAL, PARTIAL SUMS AND COMPOUND PROCESSES

Several authors have considered the impact of different forms of dependence between the severity loss random variables and its corresponding influence on the asymptotics of partial sums and compound process asymptotics in the case in which the loss random variables are from marginally heavy-tailed severity models in some sub-family of the sub-exponential heavy-tailed severity models. In this section, we consider the results, which are types of dependence structures, that one can show will result in no change to the first-order asymptotics for the compound process tails.

Geluk and Ng (2006) studied the case in which the partial sums of n loss random variables which were no longer i.i.d. where the marginal distribution of each loss $X_i \sim F_{X_i}(x)$ could differ (although practically this is not typically relevant so one may restrict to the case of identical severity distribution) and each marginal satisfied that it was a sub-member of the sub-exponential family $F_{X_i} \in \mathscr{F}$. In their study, they considered the sub-family of sub-exponential distributions given by the intersection of the dominantly varying and long-tailed distributions given by the relationship $\mathscr{D} \cap \mathscr{L} \subset \mathscr{F} \subset \mathscr{L}$. The study showed that under the assumption that the loss random variables are NA, then the first-order asymptotic relationship given in Equation 7.189 is still valid. To be precise, they studied the case of asymptotic association between

$$\mathbb{P}\mathrm{r}[R_{(n,1)} > x] \sim \mathbb{P}\mathrm{r}[\max\{X_1, \ldots, X_n\} > x] \sim \mathbb{P}\mathrm{r}[Z_n > x] \sim \sum_{i=1}^{n} \overline{F}_{X_i}(x), \quad (7.201)$$

where we recall that $R_{(n,1)}$ corresponded to the maximum of the sequence of partial sums given by $R_{(n,1)} := \max_{0 \leq i \leq n} \{Z_i\}$.

Remark 7.18 *We now start to see why it was important to recall the results for the role of the PSW identity as they aid in the solutions obtained for the analysis of the asymptotics for the dependence heavy-tailed compound processes.*

In particular, one can show that for negatively associated loss random variables with marginal distributions which have the property that they are in a sub-family of the sub-exponential family of functions formed by the intersection of the dominantly varying and long-tailed distributions $F_{X_i} \in \mathscr{D} \cap \mathscr{L} \subset \mathcal{S}$, the above asymptotic tail results apply as shown in Theorem 7.19.

Theorem 7.19 (Partial Sum Asymptotic Tail Approximations: Severity in $\mathscr{D} \cap \mathscr{L}$ and NA) *Consider the partial sums given for any $n \geq 1$ by*

$$Z_n = \sum_{i=1}^{n} X_i \sim F_{Z_n}(x), \quad (7.202)$$

with loss random variables $X_i \sim F_{X_i}$ which satisfies that the severity distribution is marginally heavy tailed with $f_X \in \mathscr{D} \cap \mathscr{L}$ and the joint severity distribution for any n given by $F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)$ satisfies for any $n \geq 1$ the conditions for negative association where for every pair of disjoint subsets A_1, A_2 of $\{1, 2, \ldots, n\}$ one has

$$\mathbb{C}\mathrm{ov}[f_1(X_i, i \in A_1),\ f_2(X_j, j \in A_2)] \leq 0, \quad (7.203)$$

whenever f_1 and f_2 increasing functions. Then if there exists a constant $c > -\infty$ such that $\mathbb{P}\mathrm{r}[X_i > c] = 1$ for all $i \in \{1, 2, \ldots, n\}$, then the following result applies for the tail asymptotic:

$$\mathbb{P}\mathrm{r}[R_{(n,1)} > x] \sim \mathbb{P}\mathrm{r}[\max\{X_1, \ldots, X_n\} > x] \sim \mathbb{P}\mathrm{r}[Z_n > x] \sim \sum_{i=1}^{n} \overline{F}_{X_i}(x), \quad (7.204)$$

where we recall that $R_{(n,1)}$ corresponded to the maximum of the sequence of partial sums given by $R_{(n,1)} := \max_{0 \leq i \leq n} \{Z_i\}$.

Remark 7.19 *In other words, when the severity distribution is the same for each loss random variable marginally and it has heavy-tailed features given by the intersection of the class of long-tailed and dominantly varying tails, then the first-order asymptotic results studied in the partial sum setting with independence are found to still hold according to*

$$\overline{F}_{Z_n}(x) \sim n\overline{F}_X(x), \quad as \;\; x \to \infty. \quad (7.205)$$

Tang (2008) studied the same class of distributions but relaxed the assumption of negative association to the weaker assumption of pairwise negative quadrant dependence. Then Ko and Tang (2008) considered partial sums of n loss random variables which were no longer i.i.d. where the marginal distribution of each loss $X_i \sim F_{X_i}(x)$ could differ and each marginal satisfied that it was a sub-member of the sub-exponential family $F_{X_i} \in \mathscr{F}$ such that the joint distribution of the loss random variables satisfied that it was dependent as quantified by a particular class of copula models. If this class of dependence models is chosen appropriately, the asymptotics for the partial sum tail $\overline{F}_{Z_n}(x)$ as $x \to \infty$ is shown to be unaffected by the dependence eventually. That is, one can show that the above asymptotic results in Equation 7.189 hold if one has a copula dependence between the loss random variables which has certain restrictions imposed on the form of copula model considered, see Cruz *et al.* (2015, chapter 8). The copula has to satisfy certain Properties, see the results in Theorem 7.13 and discussions in Albrecher *et al.* (2006) and Ko and Tang (2008). For instance, if the copula model considered can be shown to satisfy the form of negative regression dependence specified previously in 7.11, then the asymptotics are unaffected as shown in Theorem 7.20, see discussion in Ko and Tang (2008).

Theorem 7.20 (NRD via Copula Dependent Losses with Asymptotic Tail Additivity) *Consider two loss random variables X_1 and X_2 with marginal severity distributions $X_i \sim F_{X_i}$ concentrated on positive supports $[0, \infty)]$ and a joint dependence modeled by a copula distribution C that satisfies conditions of negative regression dependence for a copula (see Theorem 7.13). Then the asymptotic relationships as $x \to \infty$ apply*

$$\mathbb{P}\mathrm{r}[Z_2 > x] \sim \mathbb{P}\mathrm{r}[\max\{X_1, X_2\} > x] \sim \overline{F}_{X_1}(x) + \overline{F}_{X_2}(x), \quad (7.206)$$

under two different cases for the marginal severity tail features:

1. $F_{X_1} \in \mathscr{F}$ *and* $F_{X_2} \in \mathscr{L}$ *with* $\overline{F}_{X_2} = O(\overline{F}_{X_1})$,
2. $F_{X_i} \in \mathscr{D} \cap \mathscr{L}$ *for* $i = 1, 2$.

In OpRisk, one is typically more interested in the multivariate extension of such results for n-dependent loss random variables. In this case, one can show the following results for the partial sum tail asymptotics in Theorem 7.21 apply, for the analogous n-variate negative regression dependence definition given in Definition 7.12, see discussions in Ko and Tang (2008).

Theorem 7.21 (Multivariate NRD Losses with Asymptotic Tail Additivity) *Consider n loss random variables with marginal distributions $X_i \sim F_{X_i}$ and a joint dependence that is the analogue of the bivariate negative regression dependence for $n \geq 2$ that is captured by the following relationship on the conditional distributions of the partial sums:*

$$\frac{\mathbb{P}\mathrm{r}[Z_{j-1} > x - y | X_j = y]}{\mathbb{P}\mathrm{r}[Z_{j-1} > x - y]} = O(1), \tag{7.207}$$

which will hold uniformly for all $y \in [x_0, x]$ which should exist for some large $x_0 > 0$ such that this order of asymptotic convergence is satisfied for all $j \in \{2, \ldots, n\}$. Then one can show that the following partial sum tail asymptotics apply

$$\mathbb{P}\mathrm{r}[Z_n > x] \sim \mathbb{P}\mathrm{r}[\max\{X_1, X_2, \ldots, X_n\} > x] \sim \sum_{k=1}^{n} \overline{F}_{X_k}(x) \tag{7.208}$$

if one of the two following conditions on the marginal tail behaviour is satisfied:

1. *For all $i \in \{1, 2, \ldots, n\}$, one has $F_{X_i} \in \mathcal{F}$ and either one has that for all i, j the tail orders $\overline{F}_{X_i} = O(\overline{F}_{X_j})$ or $\overline{F}_{X_j} = O(\overline{F}_{X_i})$ for all pairs; or*
2. *For all $i \in \{1, 2, \ldots, n\}$, one has $F_{X_i} \in \mathcal{D} \cap \mathcal{L}$.*

Geluk and Tang (2009) considered the case of dependence severity models with marginal loss distributions for each of the losses in the partial sum that were from a severity distribution which was sub-exponential but unlike the previous above results, they were not considered to be in the class of dominantly varying tails. The resulting tail asymptotics detailed in Equation 7.189 hold even in the presence of multivariate dependence from say an FGM copula or other models that satisfy the dependence assumption A1 as long as the marginal severity models for the loss random variables satisfying for each marginal that $X_i \sim F_{X_i} \in \mathcal{D} \cap \mathcal{L}$. If instead the assumption A2 is satisfied (multivariate negative regression dependence), then one requires that marginal distributions satisfy $F_{X_i} \in \mathcal{F}$ for all $i \in \{1, 2, \ldots, n\}$ and for all pairs, the relationship $F_{X_i} \star F_{X_j} \in \mathcal{F}$ for all pairs $1 \leq i \neq j \leq n$.

In Albrecher *et al.* (2006), it is shown that for general Archimedean bivariate copulas, one can obtain conditions on the generator of the copula family that will guarantee the existence and validity of the partial sum of two sub-exponential loss random variables, see Proposition 7.22.

Proposition 7.22 (Partial Sum Asymptotics: Sub-exponential Archimedean Copula) *Consider two loss random variables X_1 and X_2 with marginal severity distributions that satisfy $F_{X_1} \in \mathcal{F}$ and a copula dependence given by C which is Archimedean with generator ψ that is twice differentiable and satisfies the conditions for some $M < \infty$ and $x_0 \in (0, 1)$:*

- *the first derivative of the generator is bounded at* $x = 1$ *such that* $\psi'(1) = m < 0$*, that is, negative;*
- *the second derivative of the generator is upper bounded by* $\psi''(x) \leq M$ *for all* x*; and*
- *the absolute value of the first derivative is also upper bounded* $|\psi'(x)| \leq M$ *for all* $x \in [x_0, 1]$*,*

then the following partial sum tail asymptotic applies

$$\lim_{x\to\infty} \frac{\mathbb{P}\mathrm{r}[X_1 + X_2 > x]}{\mathbb{P}\mathrm{r}[X_1 > x]} = 1 + c, \tag{7.209}$$

with $c = \lim_{x\to\infty} \overline{F}_{X_2}(x)/\overline{F}_{X_1}(x)$.

We finish this section with the results developed in Asmussen and Rojas-Nandayapa (2008) which consider the tail asymptotics for partial sums with the severity model given marginally by a LogNormal distribution and the dependence structure specified by a Gaussian copula model. The results one can derive in this case for the tail asymptotic are then given in Theorem 7.22.

Theorem 7.22 (Partial Sums with LogNormal Marginals and Gaussian Copula) *Consider a partial sum of* n *losses with marginal LogNormal distribution* $X_i \sim F_{X_i}$ *satisfying the tail asymptotic given by*

$$\overline{F}_{X_i}(x; \mu_i, \sigma_{ii}) \sim \frac{\sqrt{\sigma_{ii}}}{\sqrt{2\pi}\ln x} \exp\left(-\frac{(\ln x - \mu_i)^2}{2\sigma_{ii}}\right) \tag{7.210}$$

and a copula distribution C *given by the multivariate Gaussian copula with mean vector* $(\mu_1, \mu_2, \ldots, \mu_n)$ *and covariance matrix* $(\sigma_{ij})_{n\times n}$*. In this case, one has a partial sum tail asymptotic given by*

$$\mathbb{P}\mathrm{r}[Z_n > x] \sim m_n \overline{F}(x; \mu_*, \sigma_*), \tag{7.211}$$

with

$$\begin{aligned} \sigma_* &= \max_{1\leq k\leq n} \sigma_{kk}, \\ \mu_* &= \max_{k:\sigma_{kk}=\sigma_*} \mu_k, \\ m_n &= \text{Cardinality}\{k : \sigma_{kk} = \sigma_*, \mu_k = \mu_*\}. \end{aligned} \tag{7.212}$$

To complete the discussion on first-order asymptotics in the presence of dependence for sub-exponential marginal severity distributions, we also note the work of Foss and Richards (2010). The approach adopted involves considering multivariate loss distributions with particular conditional independence assumptions. Starting with an illustration for bivariate settings which involves the decomposition of the joint severity distribution tails for losses:

$$\begin{aligned} \mathbb{P}\mathrm{r}[X_1 + X_2 > x] &= \mathbb{P}\mathrm{r}[\max\{X_1, X_2\} > x] + \mathbb{P}\mathrm{r}[\max\{X_1, X_2\} \leq x, X_1 + X_2 > x] \\ &= \mathbb{P}\mathrm{r}[X_1 > x] + \mathbb{P}\mathrm{r}[X_2 > x] - \mathbb{P}\mathrm{r}[\min\{X_1, X_2\} > x] \\ &\quad + \mathbb{P}\mathrm{r}[\max\{X_1, X_2\} \leq x, X_1 + X_2 > x]. \end{aligned} \tag{7.213}$$

If one then assumes that marginally the losses X_i are from a severity distribution which is sub-exponential $F_{X_i} \in \mathcal{F}$, then one can state in the case that the loss random variables are independent that $\mathbb{P}\mathrm{r}[\min\{X_1, X_2\} > x]$ and $\mathbb{P}\mathrm{r}[\max\{X_1, X_2\} \leq x, X_1 + X_2 > x]$ are asymptotically negligible relative to $\mathbb{P}\mathrm{r}[X_1 > x]$. This is exactly what was used in the first-order asymptotic results above in different tail assumptions under independence, the question is when does this still occur for dependence cases. The answer to this question in the sub-exponential severity model setting has been shown above for several dependence assumptions. Here we consider a different specification of dependence through a conditioning argument that does not involve specification of particular dependence features in the form of a copula or a particular sub-class of sub-exponential distributions. The result of this analysis is provided in Proposition 7.23, see details in Foss and Richards (2010).

Proposition 7.23 (Partial Sum Asymptotics: Sub-exponential, Conditional Independence) *Consider n loss random variables $X_1, X_2, \ldots, X_n$ with marginal severity (conditional) distributions that satisfy F_{X_i} and a reference severity (unconditional) distribution $F \in \mathcal{F}$ and a dependence structure given by the following three conditional dependence assumptions:*

1. *The losses $X_1, X_2, \ldots, X_n$ are conditionally independent given a sub-sigma algebra $\mathcal{G} \subset \mathcal{T}$ of the probability space $(\Omega, \mathcal{T}, \mathbb{P})$ for which the LDA loss model is defined upon. More precisely, consider indices $\{i_1, i_2, \ldots, i_r\}$ and a collection of associated sets $\{B_{i_1}, \ldots, B_{i_r}\}$ with each $B_{i_j} \subseteq \mathcal{T}$. It is assumed that conditional on the information set $\mathcal{G}$, the probability of the loss random variables with indexes $\{i_1, i_2, \ldots, i_r\}$ satisfy the following conditional independence statement:*

$$\mathbb{P}\mathrm{r}[X_{i_1} \in B_{i_1}, X_{i_2} \in B_{i_2}, \ldots, X_{i_r} \in B_{i_r} | \mathcal{G}] = \prod_{j=1}^{r} \mathbb{P}\mathrm{r}[X_{i_j} \in B_{i_j} | \mathcal{G}]. \tag{7.214}$$

2. *The marginal distributions of each loss $X_i \sim F_{X_i}$ for all $i \in \{1, 2, \ldots, n\}$ satisfy the condition that $F_{X_i} \sim c_i \overline{F}(x)$ for constants $c_i \geq 0$ with at least one $c_i \neq 0$. In addition, for all $i \in \{1, 2, \ldots, n\}$, we will assume that there exists a c_i and a global severity distribution such that the following tail inequality holds $\overline{F}_{X_i}(x) \leq c_i \overline{F}(x)$. The condition*

$$\lim_{x \to \infty} \frac{\overline{F}_{X_i}(x)}{\overline{F}(x)} = c_i, \quad \forall i \in \{1, 2, \ldots, n\}. \tag{7.215}$$

3. *For each $i \geq 1$, there exists a non-decreasing function $r(x)$ and an increasing collection of sets $B_i(x) \in \mathcal{G}$ with $B_i(x) \to \Omega$ as $x \to \infty$ which also produce the following condition*

$$\mathbb{P}\mathrm{r}[X_i > x | \mathcal{G}] \mathbb{I}[B_i(x)] \overline{F}(x) \mathbb{I}[B_i(x) \leq r(x)], \quad a.s. \quad as \quad x \to \infty \tag{7.216}$$

such that this convergence occurs uniformly in i with the following conditions for some functions $h(x)$ and $r(x)$:

1. $\mathbb{P}\mathrm{r}[\overline{B}_i(h(x))] = o(\overline{F}(x))$;
2. $r(x)\overline{F}(h(x)) = o(1)$; *and*
3. $r(x) \int_{h(x)}^{x-h(x)} \overline{F}(x-y) F(dy) = o(\overline{F}(x))$,

where the function $h(x)$ satisfies the condition that all sub-exponential distributions $F(x) \in \mathscr{F} \subset \mathscr{L}$ are long tailed and satisfy

$$\lim_{x\to\infty} \frac{\overline{F}_X(x - h(x))}{\overline{F}_X(x)} = 1. \tag{7.217}$$

If these conditions are satisfied, then the first-order asymptotic is given by

$$\Pr[X_1 + X_2 + \cdots + X_n > x] \sim \sum_{i=1}^{n} \Pr[X_i > x] \sim \left(\sum_{i=1}^{n} c_i\right) \overline{F}_X(x). \tag{7.218}$$

Examples of LDA models and severity model distributional choices that satisfy these conditions are provided in Foss and Richards (2010).

7.6.3 SEVERITY DEPENDENCE STRUCTURES THAT DO NOT AFFECT LDA MODEL TAIL ASYMPTOTICS: CONSISTENT VARIATION

Using these concepts of negative dependence, the study of tail asymptotics for compound process LDA models was performed in Chen and Zhang (2007) where they develop large deviation results for sums of negatively associated loss random variables. In Proposition 7.24, we present the partial sum results for the case of marginal severity distributions which exhibit CV tails with the dependence given by an assumption of negative association.

Proposition 7.24 (Asymptotic Tail Approximations for Partial Sums: CV and ND) *Consider the partial sum given by*

$$Z_n = \sum_{i=1}^{n} X_i \sim F_{Z_n}(x), \tag{7.219}$$

with identically distributed loss random variables $X_i \sim f_X$ which satisfy that the severity distribution is marginally in the family $f_X \in CV$ and the joint severity distribution for any n given by $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ satisfies for any $n \geq 1$ the LND and UND conditions given by

1. LND. *A sequence of loss random variables are LND if for each $n \geq 1$ and all $X_1, X_2, \ldots, X_n$ one has*

$$\Pr[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n] \leq \prod_{i=1}^{n} \Pr[X_i < x_i]. \tag{7.220}$$

2. UND. *A sequence of loss random variables are UND if for each $n \geq 1$ and all $X_1, X_2, \ldots, X_n$ one has*

$$\Pr[X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n] \leq \prod_{i=1}^{n} \Pr[X_i > x_i]. \tag{7.221}$$

If the following condition on the severity distributions is satisfied:

- *the severity marginal distribution satisfies asymptotically as $x \to \infty$ that $x f_X(-x + \mu) = O(\overline{F}(x))$ with $\mu = \int x d f_X(x) < \infty$ and $\mathbb{E}[|X_1|^r] < \infty$ for some $r > 1$,*

then the following result applies for the tail asymptotic for each fixed $\gamma > 0$ *such that* $x > \gamma n$*:*

$$\overline{F}_{Z_n}(x) = \mathbb{P}\mathrm{r}\left[\sum_{i=1}^{n}(X_i - \mu) > x\right] \sim n\overline{F}_X(x), \quad as \;\; x \to \infty. \tag{7.222}$$

This result can then be extended to the large deviation result for the compound process under some additional assumptions on the frequency distribution tails as given in the result provided in Theorem 7.23, see Chen and Zhang (2007).

Theorem 7.23 (Asymptotic Tail Approximations for Compound Processes: CV and ND) *Consider the compound process*

$$Z_N = \sum_{i=1}^{N} X_i \sim F_{Z_N}(x), \tag{7.223}$$

with identically distributed loss random variables $X_i \sim f_X$ *which satisfies that the severity distribution is marginally consistently varying* $f_X \in CV$ *and the joint severity distribution for any* $N = n$ *given by* $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ *satisfies for any* $n \geq 1$ *the LND and UND conditions given by*

1. LND. *A sequence of loss random variables are LND if for each* $n \geq 1$ *and all* $X_1, X_2, \ldots, X_n$ *one has*

$$\mathbb{P}\mathrm{r}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n] \leq \prod_{i=1}^{n} \mathbb{P}\mathrm{r}[X_i < x_i]. \tag{7.224}$$

2. UND. *A sequence of loss random variables are UND if for each* $n \geq 1$ *and all* $X_1, X_2, \ldots, X_n$ *one has*

$$\mathbb{P}\mathrm{r}[X_1 > x_1, X_2 > x_2, \ldots, X_n > x_n] \leq \prod_{i=1}^{n} \mathbb{P}\mathrm{r}[X_i > x_i]. \tag{7.225}$$

Furthermore, the frequency distribution for the random variable on the number of losses N *is specified by* $\{p_n : n \in \mathbb{J}\}$ *which is assumed to be independent of the loss random variables. If the following conditions on the severity and frequency distributions are satisfied:*

1. *the frequency distribution satisfies that its tails are not heavy* $\mathbb{E}[N] < \infty$ *; and*
2. *The severity marginal distribution satisfies asymptotically as* $x \to \infty$ *that* $xf_X(-x + \mu) = O(\overline{F}(x))$ *with* $\mu = \int xdf_X(x) < \infty$ *and* $\mathbb{E}[|X_1|^r] < \infty$ *for some* $r > 1$*,*

then the following large deviations result applies for the tail asymptotic:

$$\overline{F}_{Z_N}(x) = \mathbb{P}\mathrm{r}\left[\sum_{i=1}^{N}(X_i - \mu) > x\right] \sim \mathbb{E}[N]\overline{F}_X(x), \quad as \;\; x \to \infty. \tag{7.226}$$

7.6.4 DEPENDENT SEVERITY MODELS: PARTIAL SUMS AND COMPOUND PROCESS SECOND-ORDER TAIL ASYMPTOTICS

To complete this discussion on dependence between loss random variables in severity models and its effect on the asymptotics on partial sums and compound processes, we note that there has been very limited study at this stage of the second-order asymptotic expansions in the case of dependence. The work of Kortschak (2012) considers the second-order expansion in the case of asymptotically tail-independent loss random variables with marginal severity distributions with regularly varying tails. As was shown in the case of independent risks above, the second-order asymptotics for regularly varying severity distribution would result in the following asymptotic expansion as $x \to \infty$:

$$\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x] = \underbrace{n\overline{F}(x)}_{\text{First-order term}} + \underbrace{\frac{n!}{(n-2)!}\mathbb{E}[X_1]f_X(x)}_{\text{Second-order term (independent)}} + O(f(x)). \tag{7.227}$$

As discussed in Kortschak (2012), the Taylor series representation of the general partial sum in the dependence case can be stated as follows:

$$\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x] = \underbrace{n\overline{F}(x)}_{\text{First-order term}} \underbrace{+(1+o(1))f_X(x)\sum_{i=1}^{n}\mathbb{E}[(Z_n - X_i)|X_i = x]}_{\text{Second-order term (dependent)}}. \tag{7.228}$$

This expression can be very difficult to evaluate because of the conditional expectations $\mathbb{E}[(Z_n - X_i)|X_i = x]$; in the presence of dependence, these expectations can be solved with numerical sampling techniques for other related problems such as consistent capital allocation and rare-event calculations, see Cruz *et al.* (2015, chapter 8).

Kortschak (2012) makes a sequence of assumptions in order to obtain analytic and closed-form approximation expressions for such expectations and, therefore, for the second-order asymptotic terms. These assumptions are listed below to demonstrate how this second-order asymptotic framework based on copulas relates to the various results presented in the first-order asymptotic dependence settings:

A.1 Assume that there exist constants $c_1 > 0$, $\varepsilon_1 \in (0, 1)$ and $\varepsilon_2 > 0$ such that for all $i \neq j$ one has

$$\lim_{a\to 0}\sup_{y\in(\varepsilon_1, 1+\varepsilon_2)}\frac{h_{i,j}(a, ya)}{h_{i,j}(a, a)} \leq c_1, \tag{7.229}$$

where the loss random variables $X_1, \ldots, X_n$ are assumed to have marginal severity distributions $f_X(x)$ with a dependence structure given by a copula distribution $C_{U_1,\ldots,U_n}(u_1, \ldots, u_n)$ where $u_i = F_{X_i}(x)$ and density given by $c_{u_1,\ldots,u_n}(u_1, \ldots, u_n)$, with the function $h_{i,j}$ defined by the marginal pairwise copula density integrated over the tail region

$$h_{i,j}(s, t) := \int_{1-s}^{1} c_{U_i,U_j}(x, (1-t))dx \tag{7.230}$$

with

$$h_{i,j}(\overline{F}_{X_i}(\delta u), \overline{F}_{X_j}(yu)) = \mathbb{P}\mathrm{r}[X_i > \delta u|X_j = uy]. \tag{7.231}$$

A.2 Assume that there exist some constants $\varepsilon_0, \gamma_1 > \gamma_2 > \frac{1}{\alpha}, c_2 > 0, c_3 > 0$ and all $(1+\varepsilon_0)(2(n-1))^\alpha < \delta < M$ uniformly for $y \in [\delta, M]$ one has the inequality for the pairwise copula function between any two loss random variables given by

$$c_2\Big(\frac{y}{\delta}\Big)^{\gamma_1} \leq \liminf_{a\to 0} \frac{h_{i,j}(ya,a)}{h_{i,j}(\delta a,a)} \leq \limsup_{a\to 0} \frac{h_{i,j}(ya,a)}{h_{i,j}(\delta a,a)} \leq c_3\Big(\frac{y}{a}\Big)^{\gamma_2}, \quad i \neq j. \tag{7.232}$$

In addition, one has the pairwise copula that satisfies that for $\varepsilon_0, \gamma_3 > \frac{1}{\alpha}, c_4 > 0$ and $\delta = (1+\varepsilon_0)(2(n-1))^\alpha$, one has uniformly for $y \in (0, \delta]$ the inequality

$$\limsup_{a\to 0} \frac{h_{i,j}(ya,a)}{h_{i,j}(\delta a,a)} \leq c_4 y^{\gamma_3}, \quad i \neq j. \tag{7.233}$$

A.3 Assume that the joint tail behaviour of the marginal distribution for losses X_i and X_j is regularly varying and, therefore, satisfies the condition

$$\lim_{u\to\infty} \frac{\mathbb{P}\mathrm{r}\,[X_i > xu, X_j > xu]}{\mathbb{P}\mathrm{r}\,[X_i > u, X_j > u]} = x^{-2\alpha/(1-\widetilde{\lambda}_{u(i,j)})} \tag{7.234}$$

with the modified upper tail coefficient $\widetilde{\lambda}_u(i,j) \in [0,1]$ given for losses X_i and X_j by

$$\widetilde{\lambda}_u(i,j) = \lim_{u\to\infty} \frac{2\ln\mathbb{P}\mathrm{r}\,[X_i > u]}{\ln\mathbb{P}\mathrm{r}\,[X_i > u, X_j > u]} - 1. \tag{7.235}$$

Using these assumptions, the following results presented in Theorem 7.24 were developed in Kortschak (2012).

Theorem 7.24 (Second-Order Asymptotics for Dependent Risks) *Consider an OpRisk loss process with n loss random variables $X_1, X_2, \ldots, X_n$ which are dependent with a copula distribution C that admits a density c and common marginal severity distribution $f_X(x)$ which is continuously differentiable and admits a density $f_X(x)$ which is regularly varying $f_X \in RV_{-\alpha-1}$. Assume also that the severity model satisfies the assumption given by A.1–A.3 and the additional conditions:*

- *Given for every pair of loss random variables X_i, X_j with $i \neq j$ that the conditional expectation $\mathbb{E}[X_i | X_j = x]$ is a function of consistent variation; and*
- *For every $\varepsilon \in \big(0, \frac{1}{2}\big]$, there exist sets $A(\varepsilon) = A(\varepsilon, m)$ such that uniformly for*

$$\left\{0 < ux < \sum_{i\neq m} u_i, \max_{i\neq m} y_i < \varepsilon u\right\} \cap \{(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n) \in A(\varepsilon)\}$$

with stochastic upper bounds given by

$$\begin{aligned}
&(1 - o_\varepsilon(1))c(F_{X_1,X_2,\ldots,X_{m-1},X_{m+1},\ldots,X_n}(\boldsymbol{u}), f_X(u)) \\
&\quad \lesssim c(F_{X_1,X_2,\ldots,X_{m-1},X_{m+1},\ldots,X_n}(\boldsymbol{u}), f_X(u(1-x))) \\
&\quad \lesssim c(F_{X_1,X_2,\ldots,X_{m-1},X_{m+1},\ldots,X_n}(\boldsymbol{u}), f_X(u(1-(n-1)\varepsilon))).
\end{aligned} \tag{7.236}$$

Then one can state the following second-order tail asymptotic for the partial sum:

$$\mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x] = n\overline{F}_X(x) + (1 + o(1))f_X(x)\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\mathbb{E}[X_j | X_i = x]. \tag{7.237}$$

Examples in the Gaussian copula setting can be found in Kortschak (2012, proposition 4.1). Next we return back to the independent case and consider higher order asymptotics for third-order and higher order tail expansions for LDA model compound process tails.

7.7 Third-order and Higher Order Single Risk Loss Process Asymptotics for Heavy-Tailed LDA Models: Independent Losses

Several approaches have been proposed in the literature to undertake higher order asymptotic expansions for the compound process distribution for the annual loss in an LDA model with heavy-tailed severity distributions. In this section, we first consider in Proposition 7.23 the third-order approximations discussed in Albrecher *et al.* (2006) which are based on the developments in Willekens (1989), Baltrūnas and Omey (1998) and Willekens and Teugels (1992) before considering two different approaches to higher order approximations which require different forms of restrictions on the severity model. The first utilizes assumptions relating to the expression of the severity distribution as a slowly varying function, whilst the second approach considers a smooth variation representation.

7.7.1 BACKGROUND UNDERSTANDING ON HIGHER ORDER TAIL DECOMPOSITION APPROACHES

In general, the majority of the approaches in the literature dealing with higher order approximations consider for a given number of losses n, a decomposition of the tail probability $\mathbb{P}\mathrm{r}[Z_n > x] = \mathbb{P}\mathrm{r}[X_1 + \cdots + X_n > x]$.

There have been several decompositions proposed in the literature, some are based on assumptions of the existence of a certain order of differentiation of the tail of the severity distribution, others are based on recursive schemes or shifted argument expansions which can avoid assumptions of differentiability.

Decomposition Approach 1. The first approach we mention involves, for a constant $w \in (0, 1)$, enumerating the possibilities related to the magnitude of the last loss X_n such that $Z_n = \sum_{i=1}^{n} X_i$ exceeds x. This is given by the following three terms:

$$\begin{aligned}\mathbb{P}\mathrm{r}[Z_n > x] = &\underbrace{\mathbb{P}\mathrm{r}[Z_{n-1} \leq (1-w)x, Z_n > x]}_{\text{Case 1: } X_n > wx} \\ &+ \underbrace{\mathbb{P}\mathrm{r}[X_n \leq wx, Z_n > x]}_{\text{Case 2: } X_n < wx, \text{ however, } Z_{n-1} + X_n > x} + \underbrace{\mathbb{P}\mathrm{r}[Z_{n-1} > (1-w)x, X_n > wx]}_{\text{Case 3: } X_n > wx \text{ and } Z_{n-1} > (1-w)x}.\end{aligned} \tag{7.238}$$

Now if one expands these terms, they can be obtained according to the following decomposition in terms of the severity distribution function,

$$\begin{aligned}\mathbb{P}\mathrm{r}[Z_n > x] &= \mathbb{P}\mathrm{r}[X_n > x - Z_{n-1} | Z_{n-1} \leq (1-w)x]\mathbb{P}\mathrm{r}[Z_{n-1} \leq (1-w)x] \\ &\quad + \mathbb{P}\mathrm{r}[Z_{n-1} > x - X_n | X_n \leq wx]\mathbb{P}\mathrm{r}[X_n \leq wx] \\ &\quad + \mathbb{P}\mathrm{r}[X_n > wx]\mathbb{P}\mathrm{r}[Z_{n-1} > (1-w)x] \\ &= \int_0^{(1-w)x} \overline{F}(x-\tau) dF^{(n-1)*}(\tau) + \int_0^{wx} \overline{F^{(n-1)*}}(x-\tau) dF(\tau) \\ &\quad + \overline{F}(wx)\overline{F^{(n-1)*}}((1-w)x). \end{aligned} \tag{7.239}$$

Decomposition Approach 2. *The second approach due to Albrecher et al.* (2006, *section 3.1) involves considering the maximum loss of the* n *losses, denoted by the order statistic* $X_{(n)} = \max_{1 \leq i \leq n} X_i$*, and then decomposing using* $X_{(n)}$ *and the difference* $Z_{(n-1)} = Z_n - X_{(n)}$ *to get*

$$\begin{aligned}&\mathbb{P}\mathrm{r}[Z_n > x] \\ &= \mathbb{P}\mathrm{r}[Z_{(n-1)} \leq x/2, X_{(n)} > x/2] + \mathbb{P}\mathrm{r}[Z_n > x, X_{(n)} \leq x/2] + \mathbb{P}\mathrm{r}[Z_n > x, Z_{(n-1)} \leq x/2].\end{aligned} \tag{7.240}$$

Note that under this approach, if the existence of derivatives of a desired order $\overline{F}_X$ *are not available, there is under additional assumptions an approach which can replace* F *with an alternative differentiable function, see Albrecher et al.* (2006, *corollary 3.6*).

Decomposition Approach 3. *The third approach also due to Albrecher et al.* (2006, *section 4) avoids the need to consider the existence of derivatives of* $\overline{F}$ *or a related function by instead formulating the following asymptotic equivalence relationship through appropriate selection of constants* $x_1, \ldots, x_m$ and $k_1, \ldots, k_m$ *such that*

$$\mathbb{P}\mathrm{r}[Z_n > x] \sim \sum_{j=1}^{m} x_j \mathbb{E}[N]\overline{F}(x - k_j). \tag{7.241}$$

7.7.2 DECOMPOSITION APPROACH 1: HIGHER ORDER TAIL APPROXIMATION VARIANTS

Having detailed these basic three types of approach considered in the literature, we first discuss how one can proceed under Decomposition Approach 1 to obtain higher order approximations. It is beneficial at this stage to consider the operator notation utilized in Barbe and McCormick (2009a, section 2) and Albrecher *et al.* (2006) when presenting this decomposition, where they define, for distribution function $F(x)$, the three operators for convolution, scaling of the argument and differentiation for a measurable test function $\phi(x)$ by

$$T_{F,w}\phi(x) = \int_0^{wx} \phi(x-\tau) dF(\tau), \text{ and}$$

$$M_c F(x) = \mathbb{P}\mathrm{r}[cX \leq x] = \begin{cases} F\left(\frac{x}{c}\right), & \text{if } c > 0, \\ \overline{F}\left(\frac{x}{c^-}\right), & \text{if } c < 0, \\ \mathbb{I}[x \geq 0], & \text{if } c = 0, \end{cases}$$

$$\Delta^r_{t,x}\phi(x) = \text{sign}(x)\frac{\phi(t(1-x)) - \phi(t)}{|x|^r\phi(t)},$$
$$\overline{\Delta}^r_{\tau,l}\phi(x) = \sup_{t\geq\tau}\sup_{0<|x|\leq l}|\Delta^r_{t,x}\phi(x)|. \tag{7.242}$$

Here the operator M_c when acting on distribution F provides the distribution for cX when c is positive, and for $c = 0$, one obtains the distribution function M_0F of the point mass at the origin. Note that for any $c > 0$, one can also observe the useful property that $\overline{M_cF}(x) = M_c\overline{F}(x)$. Furthermore, we note that the standard composition of two operators T_1 and T_2 for any smooth continuous test function $\phi(x)$ is given by $(T_1T_2)\phi = T_1(T_2\phi)$.

Having defined these operators, then certain properties of these operators, primarily involving convolutions between distribution functions, tail distribution functions and differentiation of such convolutions identities have been derived in, for example, Barbe and McCormick (2009a, section 5, proposition 5.1.1). We will recall a few of these identities in the following few propositions as they will be required for an understanding of the higher order asymptotic results.

The identities discussed in Barbe and McCormick (2009a, p. 7) will also be useful to recall, in which the tail of the convolution between two distribution functions F and G (assume for OpRisk with positive support) can be decomposed via a simple integration by parts and change of variable into a symmetrized representation according to

$$\begin{aligned}
\overline{F*G}(x) &= \int_0^\infty \overline{F}(x-\tau)\,dG(\tau) \\
&= \int_0^{(1-w)x} \overline{F}(x-\tau)\,dG(\tau) + \int_{(1-w)x}^\infty \overline{F}(x-\tau)\,dG(\tau) \\
&= \int_0^{(1-w)x} \overline{F}(x-\tau)\,dG(\tau) \\
&\quad + \left\{ \left[\overline{F}(x-\tau)(G(\tau)-1)\right]_{(1-w)x}^\infty + \int_{(1-w)x}^\infty \overline{G}(\tau)d\overline{F}(x-\tau) \right\} \\
&= \int_0^{(1-w)x} \overline{F}(x-\tau)\,dG(\tau) \\
&\quad + \left\{ \int_0^{wx} \overline{G}(x-\tau)d\overline{F}(\tau) + \overline{F}(wx)\overline{G}((1-w)x) \right\}.
\end{aligned} \tag{7.243}$$

This representation is utilized also in Barbe and McCormick (2005b, section 4, equation 4.0.2) in order to obtain an inductive decomposition of the tail distribution for the n-fold convolution $\overline{F^{(n)*}}(x)$ as presented in Proposition 7.28, where they specifically selected $w = 1/2$. First, we present the case for two distributions, which was first developed in Goldie (1978) and later considered in Baltrunas *et al.* (2006) and presented in Proposition 7.25.

Proposition 7.25 *Consider first two distribution functions F and G for independent random variables each with positive support. Then the tail distribution for the sum of these two loss random*

variables is given by the symmetrized decomposition, for any $w \in (0,1)$*, by*

$$\begin{aligned}\overline{F * G}(x) &= \overline{G}(x) + \int_0^x \overline{F}(x-\tau)\,dG(\tau) \\ &= \int_0^{(1-w)x} \overline{F}(x-\tau)\,dG(\tau) + \int_0^{wx} \overline{G}(x-\tau)\,dF(\tau) + \overline{F}(wx)\overline{G}((1-w)x) \\ &= T_{G,1-w}\overline{F}(x) + T_{F,w}\overline{G}(x) + (M_{1/w}\overline{F}M_{1/(1-w)}\overline{G})(x). \end{aligned} \tag{7.244}$$

Furthermore, if the distributions are the same $G = F$ *and* $w = 1/2$*, then one has*

$$\overline{F^{(2)*}}(x) = \overline{F}^2\left(\frac{x}{2}\right) + \int_0^{x/2} \overline{F}(x-\tau)\,dF(\tau). \tag{7.245}$$

In Barbe and McCormick (2009a, proposition 5.1.1), the following expressions for derivatives of the tail of the convolution of two distribution functions is derived with respect to the above operator specifications, presented in Proposition 7.26.

Proposition 7.26 (Differentials of Asymptotic Tails of Convolved Distributions) *Consider a distribution function* $F(x)$ *and a differentiable function* $h(x)$*, then for any* $w \in (0,1)$*, one has the following expression for the derivative of the operators* $T_{F,w}h$ *and* $M_c h$ *given by*

$$\begin{aligned} \tfrac{d}{dx}T_{F,w}h(x) &= T_{F,w}\tfrac{d}{dx}h(x) + wM_{1/(1-w)}h(x)M_{1/w}\tfrac{d}{dx}F(x), \\ \tfrac{d}{dx}M_c h(x) &= \tfrac{1}{c}M_c\tfrac{d}{dx}h(x). \end{aligned} \tag{7.246}$$

Given distribution functions F *and* G*, then for a positive integer* $k \in \mathbb{J}$*, if* F *and* G *are* k*-times differentiable, then one has the following two identities. For the distribution of the convolution between* F *and* G*, the derivative is given by*

$$\frac{d}{dx}(F * G)(x) = (F * G)^{(1)}(x) = T_{G,1-w}F^{(1)}(x) + T_{F,w}G^{(1)}(x) \tag{7.247}$$

and the k*-th order derivative for the tail distribution is given by*

$$\begin{aligned}&\frac{d^k}{dx^k}\overline{F * G}(x) = \overline{F * G}^{(k)}(x) \\ &= T_{G,1-w}\overline{F}^{(k)}(x) + T_{F,w}\overline{G}^{(k)}(x) - \sum_{1\le i\le k-1} M_{1/w}\overline{F}^{(i)}(x)M_{1/(1-w)}\overline{G}^{(k-i)}(x). \end{aligned} \tag{7.248}$$

Under further assumptions on the severity distribution F, one also obtains the recursive expression in Proposition 7.27 based on Albrecher *et al.* (2006, lemmas 3.1 and 3.3) for the case of n-losses.

Proposition 7.27 *Consider the case in which the severity distribution function has a tail distribution* $\overline{F}$ *which is regularly varying with index* ρ*,* $\overline{F} \in RV_{-\rho}$*, then if the following assumptions hold:*

1. *the severity distribution* F *has a finite fractional moment* $\mu(\gamma) = \mathbb{E}[X^\gamma] < \infty$*, for a power* $\gamma \in (0,1)$*,*

2. *for the hazard function of the severity distribution given by* $-\ln(\overline{F}(x))$*, one can find an eventually concave function* $h_0(x)$ *asymptotically equivalent to the hazard function* $h_0(x) \sim -\ln(\overline{F}(x))$*, and*
3. *for all* $\delta > 0$ *and* $c > 0$*, one has*

$$\lim_{x\to\infty} -\ln(\overline{F}(x))\overline{F}\left(\frac{cx}{-\ln(\overline{F}(x))}\right)^{\delta} = 0, \tag{7.249}$$

then one has the following expression for the tail distribution of the n*-fold convolution given for every* $\varepsilon > 0$ *by*

$$\overline{F^{(n)*}}(x) = \frac{n}{n-1}\int_0^{x/2} \overline{F^{(n-1)*}}(x-\tau)\,dF(\tau) + O\left(\overline{F}\Big(\frac{x}{2}\Big)^{2-\varepsilon}\right). \tag{7.250}$$

Under Decomposition Approach 1, the results relating to recursive decompositions and differentials can then be inductively extended to the case in which one considers the tail distribution for the sum of n i.i.d. loss random variables $Z_n = \sum_{i=1}^n X_i$ according to Proposition 7.28.

Proposition 7.28 (Recursive Decomposition of n-fold Convolution of Tail Distribution) *Given* n *i.i.d. losses the tail distribution of the resulting convolved distribution* $\overline{F^{(n)*}}(x)$ *can be decomposed as follows for any* $w \in (0,1)$ *via*

$$\begin{aligned}
\overline{F^{(n)*}} &= T_{F^{(n-1)*},1-w}\overline{F} + \left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-1)*}}\right) + T_{F,w}\overline{F^{(n-1)*}} \\
&= T_{F^{(n-1)*},1-w}\overline{F} + \left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-1)*}}\right) \\
&\quad + T_{F,w}\left[T_{F^{(n-2)*},1-w}\overline{F}(x) + \left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-2)*}}\right) + T_{F,2}\overline{F^{(n-2)*}}\right] \\
&= T_{F^{(n-1)*},1-w}\overline{F}(x) + \left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-1)*}}\right) + T_{F,w}T_{F^{(n-2)*},1-w}\overline{F}(x) \\
&\quad + T_{F,w}\left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-2)*}}\right) + T_{F,w}T_{F,w}\overline{F^{(n-2)*}} \\
&\quad \vdots \\
&= T_{F^{(n-1)*},1-w}\overline{F}(x) + \left(M_{1/w}\overline{F}M_{1/(1-w)}\overline{F^{(n-1)*}}\right) + \sum_{2\le k\le n} T_{F,w}^{k-1}T_{F^{(n-k)*},1-w}\overline{F} \\
&\quad + \sum_{2\le k\le n} T_{F,w}^{k-1}\left(M_2\overline{F^{(n-k)*}}M_2\overline{F}\right) \\
&= \sum_{1\le k\le n} T_{F,w}^{k-1}T_{F^{(n-k)*},1-w}\overline{F} + \sum_{1\le k\le n} T_{F,w}^{k-1}\left(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}\right).
\end{aligned}$$

To proceed with this recursive representation, one can either choose to approximate these integrals numerically, where one may consider importance sampling solutions, or undertake asymptotic approximations. Here the literature on higher order approximations has focussed on two different terms, resulting in two alternative higher order approximations. In Barbe and

McCormick (2004), the emphasis has been on the terms given by $T_{F^{(n-1)*},1-w}\overline{F}(x)$; however, in Albrecher *et al.* (2006), they derived an expansion by considering the terms given by $T_{F,1-w}\overline{F}^{n-1)*}(x)$.

Here we consider the representation focussed on $T_{F^{(n-1)*},1-w}\overline{F}(x)$ and we consider the approach adopted in the literature to approximate the integral operators and their derivatives by Taylor series representations and to achieve this goal, it is beneficial to utilize the operators introduced in Barbe and McCormick (2009a, section 2.1) known as the 'Laplace characters' which are given in Definition 7.22.

Before continuing the development, we note that we will make use of the representation of the tail distribution according to the hazard rate, when it exists, as given in Definition 7.20.

Definition 7.20 *Defining the hazard rate by*

$$h(x) = \frac{F^{(1)}(x)}{\overline{F}(x)}, \tag{7.251}$$

where $F^{(1)}(x)$ is the first derivative of F(x)*, and when it exists it can be used to characterize the tail of a heavy-tailed severity model according to the representation*

$$\overline{F}(x) = \overline{F}(x_0)\exp\left(-\int_{x_0}^{x} h(u)\,du\right). \tag{7.252}$$

If the hazard rate is asymptotically equivalent as follows $h(x) \sim \rho x^{-1}$, then the tail of the severity distribution is regularly varying, $\overline{F} \in RV_\rho$. If the hazard rate has a limit of $\lim_{x\to\infty} h(x) = \rho$, then $\overline{F}$ has a behaviour similar to an exponential.

Next, we consider the Taylor series expansion of the tail of a severity distribution given in Proposition 7.29, under the representation presented in Definition 7.20.

Proposition 7.29 (Taylor Series Expansion of Severity Distribution Tail) *Assume that the severity distribution tail $\overline{F}$ is differentiable up to a positive integer order $m+1 \in \mathbb{J}$, then the m-th order Taylor series representation about a point $x_0 \in [0,\infty)$ is given by*

$$\overline{F}(x) = \sum_{0\le j\le m} \frac{\overline{F}^{(j)}(x_0)}{j!}(x-x_0)^j + R_m \tag{7.253}$$

with the Lagrange remainder term given according to Abramowitz and Stegun (1972, p. 880) for some $x^* \in [x_0, x]$ by

$$R_m = \frac{(x-x_0)^{(m+1)}}{(m+1)!}\overline{F}^{m+1)}(x^*), \tag{7.254}$$

and the n-th order derivative, for all $n \le m+1$, of the tail distribution is given by Faa di Bruno's formula representation using the Bell polynomials

$$\begin{aligned}\frac{d^n}{dx^n}\overline{F}(x) = \overline{F}_n(x) &= \frac{d^n}{dx^n} f(g(x)) \\ &= \sum_{k=0}^{n} f^{(k)}(g(x)) B_{n,k}(g'(x), g''(x), \ldots, g^{(n-k+1)}(x)),\end{aligned} \tag{7.255}$$

with $f(x) = \overline{F}(x_0)\exp(-g(x))$ *and* $g(x) = \int_{x_0}^{x} h(u)du$ *and the Bell polynomials given for index* n *and* k *by*

$$B_{n,k}(x_1, x_2, \ldots, x_{n-k+1}) = \sum \frac{n!}{j_1! j_2! \cdots j_{n-k+1}!} \left(\frac{x_1}{1!}\right)^{j_1} \left(\frac{x_2}{2!}\right)^{j_2} \cdots \left(\frac{x_{n-k+1}}{(n-k+1)!}\right)^{j_{n-k+1}}, \tag{7.256}$$

where the sum is taken over all sequences $j_1, j_2, j_3, \ldots, j_{n-k+1}$ *of non-negative integers such that* $j_1 + j_2 + \cdots = k$ *and* $j_1 + 2j_2 + 3j_3 + \cdots = n$.

It is instructive to present a few examples of terms present in the series expansion,

$$\begin{aligned}
\overline{F}^{(1)}(x) &= -h(x)\overline{F}(x), \\
\overline{F}^{(2)}(x) &= (-h^{(1)}(x) + h^2(x))\overline{F}(x), \\
\overline{F}^{(3)}(x) &= (-h^{(2)}(x) + 3h^{(1)}(x)h(x) - h^3(x))\overline{F}(x).
\end{aligned}$$

In general, one can prove, see Barbe and McCormick (2005b, lemma 4.1.1), that the k-th derivative of the tail of a sub-exponential severity distribution with hazard rate which is regularly varying and satisfying

$$\lim_{x\to\infty} xh(x) = \infty, \text{ and } \lim_{x\to\infty} h(x) = 0 \tag{7.257}$$

is of asymptotic order

$$\overline{F}^{k)} \sim (-1)^k h^k \overline{F}, \tag{7.258}$$

which is beneficial in asymptotic approximations for deciding the order of remainder terms in the truncation of the series representation. The substitution of this series expansion for $\overline{F}$ into the recursive decomposition requires that we consider terms discussed in Proposition 7.30.

Proposition 7.30 (Series Expansion of Decomposition of n-Fold Convolution of the Tail) *Given* n *i.i.d. losses, the tail distribution of the resulting convolved distribution* $\overline{F^{(n)*}}(x)$ *is decomposed for any* $w \in (0,1)$ *via*

$$\overline{F^{(n)*}}(x) = \sum_{1\le k\le n} T_{F,w}^{k-1} T_{F^{(n-k)*},1-w}\overline{F}(x) + \sum_{1\le k\le n} T_{F,w}^{k-1}\left(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}\right)(x).$$

Assuming that the severity distribution tail $\overline{F}$ *is differentiable up to a positive integer order* $m+1 \in \mathbb{J}$, *then the* m*th order Taylor series representation about a point* x_0 *is given by considering the approximation terms and remainder terms substituted into this recursions. This will produce for each fixed* $k \in \{1, \ldots, n\}$ *and a given order of approximation* m *and for a value of*

$\kappa \in \{w, (1-w)\}$ *the terms*

$$T_{F^{(n-k)*},\kappa}\overline{F}(x) \approx \underbrace{\sum_{0\le j\le m} \frac{\overline{F}^{(j)}(x_0)}{j!} T_{F^{(n-k)*},\kappa}(x-x_0)^j}_{m\text{th order truncated Taylor approximation}}$$
$$= \sum_{0\le j\le m} \frac{\overline{F}^{(j)}(x_0)}{j!} \int_0^{\kappa x} (x-x_0-\tau)^j dF^{(n-k)*}(\tau)$$
$$= \sum_{0\le j\le m} \frac{(-1)^j \overline{F}^{(j)}(x)}{j!} \int_0^{\kappa x} \tau^j dF^{(n-k)*}(\tau), \text{ for } x = x_0 \in [0,\infty),$$

and the remainder term for any $x_* \in [x_0, x]$ *given as a result of the linear properties of the operator* $T_{F^{(n-k)*},\kappa}$ *by*

$$T_{F^{(n-k)*},\kappa}R_m(x) = \frac{\overline{F}^{(m+1)}(x^*)}{(m+1)!} T_{F^{(n-k)*},\kappa}(x-x_0)^{m+1}$$
$$= \frac{\overline{F}^{(m+1)}(x^*)}{(m+1)!} \int_0^{\kappa x} (x-x_0-\tau)^{m+1} dF^{(n-k)*}(\tau)$$
$$= \frac{(-1)^{m+1}\overline{F}^{(m+1)}(x)}{(m+1)!} \int_0^{\kappa x} \tau^{m+1} dF^{(n-k)*}(\tau), \text{ for } x = x_0 \in [0,\infty).$$

From the second set of recursive terms, one has the following expressions for each $k \in \{1, \ldots, n\}$ *given by*

$$T_{F,w}^{k-1}\left(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overrightarrow{\overline{F}}\right)(x)$$
$$= T_{F,w}^{k-2} \int_0^{wx} \overline{F^{(n-k)*}}((1-w)x-\tau_1)\overline{F}(wx-\tau_1)dF(\tau_1). \tag{7.259}$$

The remainder terms can be obtained analogously for this second set of terms by substitution of the series expansion remainder for $\overline{F}$ *into each term.*

One can also characterize the order of approximation for a given Taylor series truncation point m according to the results in Proposition 7.31. If one desires asymptotic approximations, these order of approximation arguments can then be used to justify asymptotically ($x \to \infty$) discarding several terms in the above decomposition as discussed below. It will first be beneficial to define as in Barbe and McCormick (2005b) the pseudo-semi-norm on tails of severity distributions given by Definition 7.21.

Definition 7.21 *For given distributions* K *and* F*, define the pseudo-semi-norm for the tail of the distribution according to*

$$|\overline{K}|_F := \sup_{x\ge 0} \frac{\overline{K}(x)}{\overline{F}(x)} \tag{7.260}$$

with convention $0/0 = 0$*. This pseudo-semi-norm for distribution tails then generates balls* $B(F,\lambda)$ *containing all tails* $\overline{K}$ *which are less than* $\lambda\overline{F}$ *and* $B(F)$ *denotes the union of all such balls for a positive* $\lambda > 0$*.*

This can then be utilized to make the following statements about the asymptotic order of the terms in the tail decomposition as follows.

Proposition 7.31 (Order of Approximation for Taylor Series for the *n*-Fold Tail Convolution) *Consider the Taylor series approximation of the terms comprising the recursive decomposition of the n-fold tail convolution in Proposition 7.28. Then in the cases of severity distributions in which the tail distribution is regularly varying with $\overline{F}(x) \in RV_{-\rho}$ in which for all $\lambda > 0$, one has*

$$\lim_{x\to\infty} \frac{\overline{F}(x\lambda)}{\overline{F}(x)} = x^{-\rho}. \tag{7.261}$$

Albrecher et al. *(2006) stated that for $R_m(x) \sim O\left(\overline{F}^{(m)}(x)\right)$, the term $T_{F^{(n-k)*},\kappa}R_m(x)$ is of the same asymptotic order of approximation for all $k \leq n$, that is, $T_{F^{(n-k)*},\kappa}R_m(x) \sim O\left(\overline{F}^{(m)}(x)\right)$. In addition, under this assumption, for all $j \leq m$ and $k \leq n$, the terms given by*

$$\overline{F}^{(j)}(x)\int_{\tau x}^{\infty} \tau^j dF^{(k)*}(\tau) = O\left(\overline{F}^{m)}(x)\right) \tag{7.262}$$

have the same order of approximation as the remainder terms, because of the assumption of a monotone density and Karamata's theorem. Furthermore, in Barbe and McCormick (2005b, lemma 4.2.2), it was shown that assuming the hazard rate $h(x)$ exists, for sub-exponential distribution F, and is regularly varying with negative index, then for a distribution K in the ball $B(F,\lambda)$ and any function g of order of approximation $O\left(h^m\overline{F}\right)$,

$$T_{K,w}g = O\left(h^m\overline{F}\right). \tag{7.263}$$

Then from the second set of terms in the decomposition, the asymptotic behaviour for each of the terms for $k \in \{1, 2, \ldots, n\}$ was proven in Barbe et al. *(2007) that there exists an interval $[t,\infty)$, a positive number ε and constant A such that for any positive $k \leq n$*

$$\left|T_{F,w}^{k-1}\left(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}\right)\right| \leq A(1+\varepsilon)^n h^{m+1}\overline{F}. \tag{7.264}$$

One can then proceed in a manner of different ways: approximation outright of the integrals involved in the decomposition numerically (discussed in the section on importance sampling); using the knowledge of the orders of approximation of particular terms, approximation of the remaining dominant terms using numerical procedures; or utilizing the orders of approximation to find a closed-form expression that is asymptotically valid. This third approach will be discussed here and the first two alternatives are deferred to the importance sampling approach presented later. Next we present the resulting asymptotic Taylor series expansion which is modified utilizing terms of the same order of approximation to simplify the expressions, as shown in Proposition 7.32.

Proposition 7.32 (Modified Taylor Series Expansion of n-fold Tail Convolution) *Consider the Taylor series approximation of the decomposition of the convolved distribution $\overline{F^{(n)*}}(x)$ for any $w \in (0,1)$ and assuming that the severity distribution tail $\overline{F}$ is differentiable up to a positive integer order $m+1 \in \mathbb{J}$, then the mth order Taylor series representation about a point $x = x_0$ is*

given asymptotically in $x \to \infty$ up to order of approximation $O\left(\overline{F}^{m}\right)(x))$ by the modified Taylor series

$$\begin{aligned}
\overline{F^{(n)*}}(x) &= \sum_{1\le k\le n} T_{F,w}^{k-1} T_{F^{(n-k)*},1-w}\overline{F}(x) + \sum_{1\le k\le n} T_{F,w}^{k-1}\left(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}\right)(x) \\
&\approx \sum_{1\le k\le n} T_{F,w}^{k-1}\left[\sum_{0\le j\le m}\frac{(-1)^j\overline{F}^{(j)}(x)}{j!}\left(\int_0^{\kappa x}\tau^j dF^{(n-k)*}(\tau)\right.\right. \\
&\qquad \left.\left. + \int_{\kappa x}^{\infty}\tau^j dF^{(n-k)*}(\tau)\right)\right] \\
&= \sum_{1\le k\le n} T_{F,w}^{k-1}\left[\sum_{0\le j\le m}\frac{(-1)^j\overline{F}^{(j)}(x)}{j!}\int_0^{\infty}\tau^j dF^{(n-k)*}(\tau)\right] \\
&= \sum_{1\le k\le n}\sum_{0\le j\le m} T_{F,w}^{k-1}\frac{(-1)^j\mu_{F^{(n-k)*},j}}{j!}\overline{F}^{(j)}(x).
\end{aligned}$$

To proceed, it becomes beneficial to introduce the notion of the 'Laplace character' as provided in Definition 7.22, see Barbe and McCormick (2004). This is primarily due to the relationship this operator has in converting compositions with respect to a distribution into a single operator with respect to a convolution between the distributions. A detailed discussion of this property and the motivation for Laplace characters is best found in Barbe and McCormick (2004, sections 1 and 2).

Definition 7.22 (Space of Linear Differential Operators and Laplace Characters) *Consider the differential operator D^k operating on a test function $\phi(x)$ which is m-th order differentiable $\phi \in \mathbb{C}^m[\mathbb{R}]$ with $k \le m$ and given by $D^k\phi(x) = \frac{d^k}{dx^k}\phi(x)$, defining D^0 as identity. Then consider the linear differential operator given by $\sum_{0\le i\le m} p_i D^i$ and defined over the space of all such real linear differential operators with constant coefficients and order at most m, denoted by $\mathbb{R}_m[D]$. On this space, one can then define a special composition operation between two elements of the space, p and q, given by $p \circ q$ according to*

$$\begin{aligned}
p &= \sum_{0\le i\le m} p_i D^i, \text{ and } q = \sum_{0\le i\le m} q_i D^i, \\
p\circ q &= \sum_{0\le i\le m}\left(\sum_{0\le j\le i} p_i q_{i-j}\right) D^i.
\end{aligned} \tag{7.265}$$

Now given a severity distribution F, where for some positive $m \in \mathbb{J}$, the m-th absolute moment with respect to F exists and is finite, denoted for any $0 \le k \le m$ by $\mu_{F,k} = \int x^k dF(x)$. Then the m-th 'Laplace Character' is given by

$$L_{F,m} = \sum_{0\le k\le m}\frac{(-1)^k}{k!}\mu_{F,k}D^k. \tag{7.266}$$

Note that this composition property on the space $\mathbb{R}_m[D]$ between two differential operators with $i + j > m$ given by $D^i \circ D_j = 0$. Now consider two distribution functions F and G, the Laplace

character operators have the following composition property with respect to convolutions of distribution functions

$$L_{F,m} \circ L_{G,m} = L_{F*G,m} = \sum_{0 \leq k \leq m} \frac{(-1)^k}{k!} \mu_{F*G,k} D^k = \sum_{0 \leq k \leq m} \frac{(-1)^k}{k!} \mu_{G,k} L_{F,m-j} D^k. \tag{7.267}$$

Additionally, the following property holds:

$$L_{M_c F,m} M_c = M_c L_{F,m}. \tag{7.268}$$

It will also be beneficial to observe the result of Barbe and McCormick (2009a, proposition 2.1.2) which provides an expression for the inverse of the Laplace characters, as presented in Proposition 7.33.

Proposition 7.33 (Inverse Laplace Characters) *Define the set of all ordered partitions of n which are of length at most m by $\mathscr{P}(m,n)$, where an element of this set $p \in \mathscr{P}(m,n)$ is the m-tuple of positive integers $p = (p_1, \ldots, p_m)$ such that $p_1 \geq p_2 \geq \cdots \geq p_m > 0$ and $\sum_{k=1}^m p_k = n$. Additionally, define the difference operator $\Delta p = (p_1 - p_2, p_2 - p_3, \ldots, p_m - p_{m+1})$, and for a tuple $p \in \mathscr{P}(m,n)$ and an integer k, define the multinomial-type coefficient*

$$\binom{k}{p} = \binom{k}{p_1, \ldots, p_m} = \frac{k!}{p_1! p_2! \cdots p_m!}. \tag{7.269}$$

In the space $\mathbb{R}_m[D]$, the following inverse of the Laplace character is defined for a distribution F,

$$L_{F,m}^{-1} = \sum_{0 \leq n \leq m} \left(\sum_{p \in \mathscr{P}(m,n)} (-1)^{n+p_1} \binom{p_1}{\Delta p} \prod_{1 \leq k \leq m} \left(\frac{\mu_{F,k}}{k!} \right)^{(\Delta p)_k} \right) D^n. \tag{7.270}$$

These Taylor polynomial operators are used to asymptotically approximate the Laplace transform which will generally not exist for heavy-tailed distributions, i.e. do not admit the required integrability. Put another way, these operators are the m-th order Taylor polynomials of the Laplace transform of distribution F, with the variable of F replaced with the differential operator D. That is, we expand the Laplace transform of the distribution F up to mth order at D.

So now we need to convert the representation we have obtained so far for the modified Taylor approximations of the decomposition of the n-fold convolution up to order $O\left(\overline{F}^{m)}\right)$ into a representation based on the properties of the Laplace characters. To achieve this goal, one must study how the integral operators $T_{F,w}\phi(x)$ are approximated by the Laplace characters. This was first addressed in Barbe and McCormick (2009a, section 5) owing to the result presented in Proposition 7.34. It is also relevant to recall the definition of a smoothly varying function in Definition 3.20. Using the definition of smoothly varying functions, one can bound the difference between integral operators for a distribution F given by $T_{F,w}$ and the Laplace characters over such functions according to the result in Proposition 7.34.

Proposition 7.34 (Approximating Integral Operators with Laplace Characters) *Consider a severity distribution F with positive support and finite m-th moment for some positive integer*

$m \in \mathbb{J}$. *Given constants* $r \in [0,1)$ *and* $w \in (0,1)$ *and a smoothly varying function* h *of order* $m+r$, *then the following bound holds:*

$$|(T_{F,w} - L_{F,m})h|(x) \leq \frac{|h^{(m)}(x)|}{m!x^r} \int_0^{wx} \overline{\Delta}^r_{x,\tau/x}(h^{(m)}) x^{m+r}\, dF(x) + \sum_{0 \leq j \leq m} \frac{|h^{(j)}(x)|}{j!} \int_{wx}^{\infty} x^j\, dF(x). \tag{7.271}$$

Using this results recursively, it was shown in Barbe and McCormick (2004) and Barbe and McCormick (2005b) that the Taylor series decomposition in Proposition 7.31 can be converted into a representation in terms of the Laplace characters according to Proposition 7.35.

Proposition 7.35 (Laplace Characters Taylor Expansion of n-fold Tail Convolution) *Consider the Taylor series approximation of the decomposition of the convolved distribution* $\overline{F^{(n)*}}(x)$ *for any* $w \in (0,1)$ *and assuming that the severity distribution tail* $\overline{F}$ *is differentiable up to a positive integer order* $m+1 \in \mathbb{J}$, *then the* m*-th order Taylor series representation about a point* $x = x_0$ *is given asymptotically in* $x \to \infty$ *up to order of approximation* $O(\overline{F}^{(m)}(x))$ *by the modified Taylor series*

$$\begin{aligned}
\overline{F^{(n)*}}(x) &= \sum_{1 \leq k \leq n} \sum_{0 \leq j \leq m} T^{k-1}_{F,w} \frac{(-1)^j \mu_{F^{(n-k)*},j}}{j!} \overline{F}^{j)}(x) + O\left(\overline{F}^{(m)}(x)\right) \\
&= \sum_{1 \leq k \leq n} L^{k-1}_{F,m} \circ L_{F^{(n-k)*},m} \overline{F}(x) + O\left(\overline{F}^{(m)}(x)\right) \\
&= \sum_{1 \leq k \leq n} \underbrace{L_{F^{(k-1)*} * F^{(n-k)*},m}}_{\text{Note the } k\text{-th term is deleted}} \overline{F}(x) + O\left(\overline{F}^{(m)}(x)\right) \\
&= n L_{F^{(n-1)*},m} \overline{F}(x) + O\left(\overline{F}^{(m)}(x)\right), \text{ if } \forall i \in \{1, \ldots, n\}, X_i \overset{i.i.d.}{\sim} F(x).
\end{aligned}$$

This result can be extended to the case of an infinitely weighted sum of i.i.d. random variables $c_i X_i$ for some sequence of real constants $(c_i)_{i \in \mathbb{J}^+}$ which results in consideration of distributions for the weighted sum $\sum_{i=1}^n c_i X_i$ with distribution $F_c^{(n)*}$. The higher order asymptotics of the tail distribution for such a setting was the main result of Barbe and McCormick (2004, theorem 2.3.1) for the case of distribution F being heavy tailed and sub-exponential, as discussed in Theorem 7.25.

Theorem 7.25 (Laplace Characters Taylor Expansion of Infinitely Weighted Tail Convolutions) *Given a* w *at least 1. Let* F *be a distribution function which has a tail* $\overline{F}$ *which is smoothly varying with index* $-\rho$ *and order* p, $\overline{F} \in SR_{-\rho,p}$ *because it is ultimately, as* $x \to \infty$, p *times continuously differentiable and* $\overline{F}^{(p)} \in RV_{-\rho-p}$. *Given two integers* $m < \rho$ *and* k *with* $m + k < p$ *and a positive sequence of constants* $c = (c_i)_{i \in \mathbb{J}}$ *and let* $\gamma \in (0, p - m - k \vee 1)$, *then there exists a function* $\nu(\cdot)$ *converging to 0 at infinity and a real number* x_0 *such that for any* $x > x_0$, *for any sequence* $(c_i)_{i \in \mathbb{J}}$ *which satisfies that* $N_{\rho,\gamma,p} < \infty$, *where*

$$N_{\rho,\gamma,p}(c) = |c|_{\gamma\left(\frac{\rho}{\rho+p} \wedge \frac{1}{2}\right)} \vee 2^{\frac{\rho}{\rho+p}} |c|_\infty \tag{7.272}$$

with respect to the l_p norm of a sequence $|c|_p = \left(\sum_{i\in\mathbb{J}}|c_i|^p\right)^{1/p}$. Then the following asymptotic equivalence for the tail expansion of the tail of the distribution $\sum_{i\in\mathbb{J}}c_iX_i \sim G_c(x)$, with i.i.d. $X_i \sim F$, applies

$$\overline{G}_c(x)^{(k)} \sim \sum_{i\in\mathbb{J}} L_{G_c\natural M_{c_i}F,m}(\overline{M_{c_i}F})^k(x), \tag{7.273}$$

where the operator $\natural$ denotes deletion of the ith term from the convolution $G_c\natural M_{c_i}F$, indicating the distribution for $\sum_{i\in\mathbb{J}\setminus\{i\}}c_iX_i \sim (G_c\natural M_{c_i}F)(x)$.

This result was also studied in the case of an infinitely weighted sum of i.i.d. random variables in the case in which the severity distributions F is considered to be a light sub-exponential distribution, lighter than regularly varying in which the distributions considered satisfy the hazard rate conditions that

$$\lim_{x\to\infty} xh(x) = \infty, \quad \text{and} \quad \lim_{x\to\infty} h(x) = 0. \tag{7.274}$$

This resulted in the asymptotic higher order series expansion of the form given in Barbe and McCormick (2005b, theorem 2.3.1) for hazard functions satisfying that $h(x) >> x^{-1}\ln(x)$ and with $h(x)$ smoothly varying with negative index and an order of m. If one considers the tail distribution of $\sum_{i\in\mathbb{J}}c_iX_i$ for a sequence of positive weights $c = (c_i)_{i\in\mathbb{J}}$, decreasing with index i with $c_1 = 1$, then for any $k < m$, one has the expansion

$$\overline{G} = \overline{F} - \mu_{H,1}\overline{F}^{(1)} + \frac{\mu_{H,2}}{2}\overline{F}^{(2)} + \cdots + (-1)^k\frac{\mu_{H,k}}{k!}\overline{F}^{(k)} + O\left(h^{(k)}\overline{F}\right). \tag{7.275}$$

Especially, for the case of OpRisk modeling, this was illustrated for the case of a compound process such as observed in an LDA model framework with the result presented in Barbe *et al.* (2007, theorem 2.1) given in Theorem 7.26.

Theorem 7.26 (Higher Order Tail Approximations for LDA Light Sub-exponential Models) *Given a severity distribution F with a hazard rate h that is smoothly varying with index $\rho \in [-1, 0)$ and positive order m. Furthermore, assume that the hazard rate satisfies*

$$\liminf_{x\to\infty} \frac{xh(x)}{\ln x} > 0, \tag{7.276}$$

and the moment-generating function of the frequency distribution with finite moment-generating function in the neighbourhood of the origin, then for any integer $k \leq m$,

$$\overline{F}_{Z_N} = \mathbb{E}[N]L_{F^{(N-1)*},k}\overline{F} + O\left(h^k\overline{F}\right). \tag{7.277}$$

7.7.3 DECOMPOSITION APPROACH 2: HIGHER ORDER TAIL APPROXIMATIONS

In this section, we briefly detail the alternative higher order approximations considered in Albrecher *et al.* (2006) which were denoted in this chapter as Decompositions 2 and 3. The following lemmas derived, presented in Albrecher *et al.* (2006), are required for an understanding of the results one derive under the alternative Decompositions 2 and 3.

Lemma 7.7 *If the tail distribution $\overline{F}$ is regularly varying with index $\rho > 0$, $\overline{F} \in RV_{-\rho}$, then for two or more losses, $n \geq 2$, one has the following relationship between the joint distribution of the sum of the n losses Z_n and the maximum of the losses $X_{(n)}$ given by*

$$\Pr[Z_n > x, X_{(n)} \leq x/2] = O(\overline{F}(x)^2). \tag{7.278}$$

Secondly, one can show that if the tail distribution is smoothly varying with index ρ and order 2, then the result in Lemma 7.8 holds.

Lemma 7.8 *If the tail distribution $\overline{F}$ is smoothly varying with index $\rho > 2$ and order 2, $\overline{F} \in SR_{-\rho,2}$, then for two or more losses, $n \geq 2$, one has the following asymptotic equivalence relationship between the joint distribution of the sum of the n losses Z_n and the maximum of the losses $X_{(n)}$ given by*

$$\Pr[Z_n > x, X_{(n)} \leq x/2] \sim \binom{n}{2} \mathbb{E}[Z_{n-2}^2] f(x/2)^2, \tag{7.279}$$

where $f(x)$ is the density corresponding to severity distribution F.

If the tail of the severity distribution is not regularly varying, it was shown in Albrecher *et al.* (2006, lemma 3.3) that one could still find an order of approximation for the joint probability $\Pr[Z_n > x, X_{(n)} \leq x/2] = O(\overline{F}(x/2)^{2-\varepsilon})$ as specified in Lemma 7.9.

Lemma 7.9 *If the following assumptions hold:*

1. *the distribution $F(x)$ has some finite power mean $\mu_\gamma = \mathbb{E}[X^\gamma] < \infty$ for some $0 < \gamma < 1$;*
2. *for the hazard function $-\ln(\overline{F}(x))$, one can find an eventually asymptotically equivalent concave function $h_0(x)$ with $h_0(x) \sim \ln(\overline{F}(x))$; and*
3. *for all $\delta > 0$ and $c > 0$, the asymptotic limit $\lim_{x\to\infty} -\ln(\overline{F}(x))\overline{F}\left(\frac{cx}{-\ln(\overline{F}(x))}\right)^\delta = 0$,*

then for all $0 < \varepsilon < 1$ and $K > 1$, one can find a constant $M = M(\varepsilon, K) > 0$ such that the following bound holds for all $x > M$ and $n \geq 2$, according to

$$\Pr[Z_n > Kx, X_{(n)} \leq x] \leq (1+\varepsilon)^n \overline{F}(x)^{K(1+\varepsilon)}. \tag{7.280}$$

Then using the definition of smoothly varying functions and these properties in Lemmas 7.7 and 7.8 as the alternative decompositions, it was shown in Barbe and McCormick (2004) that for the class of smoothly varying severity distribution functions, the following result for the compound distribution tail $\overline{F}_{Z_N}$ is attainable. If one considers severity distribution tails $\overline{F}(x)$ which are smoothly varying with index ρ and order $u = m + r$ for some $m \in \mathbb{N}$ and $r \in [0, 1)$ such that for $k < \min(\rho + 1, u + 1)$, one has the tail of the compound distribution represented by $\overline{F}_{Z_n} = a_k(x) + O(\overline{F}^{k-1)}(x))$, where $a_k(x)$ is given by

$$a_k(x) = a_1(x) + \sum_{j=0}^{k-2} \frac{(-1)^j \mathbb{E}[N(X_1 + \cdots + X_{N-1})^{j-1}]}{(j+1)!} f^{(j)}(x). \tag{7.281}$$

This result is then utilized in a result derived by Albrecher *et al.* (2006, lemma 2.4) which is given in Lemma 7.10 and utilized to derive an alternative higher order representation under

Decomposition 2. Recall that under Decomposition 2, one considers the tail representation of the n-fold convolution of the distributions given by

$$\begin{aligned}\mathbb{P}\mathrm{r}[Z_n > x] = \mathbb{P}\mathrm{r}[Z_{(n-1)} \leq x/2, X_{(n)} > x/2] + \mathbb{P}\mathrm{r}[Z_n > x, X_{(n)} \leq x/2] \\ + \mathbb{P}\mathrm{r}[Z_n > x, Z_{(n-1)} \leq x/2].\end{aligned}$$

Then one can find the asymptotic equivalence between the third term in this expansion and the function $a_k(x)$ as detailed in Lemma 7.10.

Lemma 7.10 *Consider a tail distribution that is smoothly varying with index $-\rho$ and order $\lceil \rho \rceil$, $\overline{F}(x) \in SR_{-\rho,\lceil \rho \rceil}$. If $\rho \neq k+1$, then the term $\mathbb{P}\mathrm{r}[Z_n > x, Z_{(n-1)} \leq x/2]$ in the tail decomposition has the following asymptotic order of approximation with respect to $a_k(x)$ in Equation 7.281 given by*

$$\mathbb{P}\mathrm{r}[Z_n > x, Z_{(n-1)} \leq x/2] - a_{k+1}(x) \sim -\binom{n}{2} \overline{F}(x)^2((1-2\rho)B(1-\rho, 1-\rho) + 2^{2\rho}),$$

where $B(a,b)$ is the Beta function. Furthermore, if $\rho = k+1$, then one has asymptotic equivalence given by

$$\mathbb{P}\mathrm{r}[Z_n > x, Z_{(n-1)} \leq x/2] - a_{k+1}(x) \sim \frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \int_0^{x/2} x^{k+1} dF^{(n-1)*}(x).$$

Furthermore, in Albrecher *et al.* (2006), note that for any integer $k > 0$, if the tail distribution of a smoothly varying distribution has index $-\rho(k+1)$ and order $k+1$, $\overline{F}(x) \in SR_{-(k+1),k+1}$ and the mean is finite $\mathbb{E}[X_1] < \infty$, then the following asymptotic equivalence can be shown,

$$\begin{aligned}&\frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \int_0^{x/2} \tau^{k+1} dF^{(n-1)*}(\tau) \\ &\sim \frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \left(\int_0^{x/2} \tau^{k+1} dF^{(n-1)*}(\tau) + \int_{x/2}^{\infty} \tau^{k+1} dF^{(n-1)*}(\tau) \right) \\ &\sim \frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \mathbb{E}\left[Z_{n-1}^{k+1}\right]. \end{aligned} \tag{7.282}$$

However, if the severity distribution does not have a finite mean, $\mathbb{E}[X_1] = \infty$, then one has the asymptotic equivalence result given by

$$\begin{aligned}&\frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \int_0^{x/2} \tau^{k+1} dF^{(n-1)*}(\tau) \\ &\sim \frac{n(n-1)(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \int_0^{x/2} \tau^{k+1} dF(\tau). \end{aligned} \tag{7.283}$$

From these results which rely on the existence of $k+1$ continuous derivatives for the tail distribution $\overline{F}$, Albrecher *et al.* (2006) showed that one then obtains the following higher order approximation of the tail distribution of an n-fold convolution given in Theorem 7.27.

Theorem 7.27 (Tail Approximations of Eventually $k+1$ Differentiable Distributions) *Consider n i.i.d. losses $X_1, \ldots, X_n$ with common severity distribution with tail that is smoothly varying with index $-\rho$ and order $\lceil \rho \rceil$, $\overline{F}(x) \in SR_{-\rho,\lceil \rho \rceil}$. If $k < \rho < k+1$, then*

$$\Pr[Z_n > x] = a_{k+1}(x) - \binom{n}{2} \overline{F}(x)^2 (1-2\rho) B(1-\rho, 1-\rho) + O(\overline{F}(x)^2) \tag{7.284}$$

otherwise, if $\rho = k+1$, one has

$$\begin{aligned} \Pr[Z_n > x] = a_{k+1}(x) + \frac{n(-1)^{k+1}\overline{F}^{(k+1)}(x)}{(k+1)!} \int_0^{x/2} \tau^{k+1} dF^{(n-1)*}(\tau) \\ + o\left(\overline{F}^{(k+1)}(x) \int_0^{x/2} \tau^{k+1} dF^{(n-1)*}(\tau) \right). \end{aligned} \tag{7.285}$$

In the case in which one cannot assume that the tail distribution is regularly varying, then Albrecher *et al.* (2006) demonstrated how one could utilize an auxiliary function in the series expansion which does have $k+1$th order derivatives eventually existing. This relied on the existence of such a function as long as the absolute difference between the severity distribution $F(x)$ and the auxiliary distribution function $H(x)$ is regularly varying, see details in Albrecher *et al.* (2006, corollary 3.6) and Geluk *et al.* (2000).

Previously, the terms $T_{F^{(n-1)*},1-w}\overline{F}(x)$ were considered to derive series representations. In the following, we note the result obtained in Albrecher *et al.* (2006, lemma 3.7) which instead focusses on the term $T_{F,1-w}\overline{F}^{(n-1)*}(x)$ with $w = 1/2$ and is given in Lemma 7.11.

Lemma 7.11 *For all $x > 0$ and number of losses $n \geq 2$, there exists functions $\overline{R}_1$ and $\overline{R}_2$ with bounds*

$$\begin{aligned} \overline{R}_1(x) &\leq \Pr[X_1 > x/2]^2 = \overline{F}(x/2)^2, \\ \overline{R}_2(x) &\leq \Pr[Z_n > x, X_{(n)} \leq x/2], \end{aligned} \tag{7.286}$$

such that the following relationship holds:

$$\begin{aligned} \Pr[Z_n > x] &= \frac{n}{n-1} \int_0^{x/2} \Pr[Z_{n-1} > x - \tau] dF(\tau) + n(\overline{R}_1(x) + \overline{R}_2(x)) \\ &\leq \frac{n}{n-1} \int_0^{x/2} \Pr[Z_{n-1} > x - \tau] dF(\tau) \\ &\quad + n(\overline{F}(x/2)^2 + \Pr[Z_n > x, X_{(n)} \leq x/2]). \end{aligned} \tag{7.287}$$

Finally, one can also derive results under Decomposition 3, which avoid the need to evaluate derivatives of the tail distribution. This expansion utilizes the results presented in Decompositions 1 and 2 and so the details are omitted, see Albrecher *et al.* (2006, section 4) for details.

7.7.4 EXPLICIT EXPRESSIONS FOR HIGHER ORDER RECURSIVE TAIL DECOMPOSITIONS UNDER DIFFERENT ASSUMPTIONS ON SEVERITY DISTRIBUTION BEHAVIOUR

In this section we develop expressions for higher order asymptotic approximations of the compound process tails, particularly considering 3-rd order approximations.

Definition 7.23 (Third-Order Single-Loss Approximation) *If the severity distribution function is regularly varying $F \in RV_{-\rho}$ with index $\rho > 2$ and one of the two following assumptions on the severity density also holds:*

1. *the derivative of the density function is regularly varying such that $df/dx \in RV_\sigma$ for some σ; or*
2. *the second derivative of the severity density d^2f/dx^2 that exists is asymptotically decreasing and there exists a regularly varying function $t(x) \in RV_\sigma$ for $\sigma \in (0, 1]$ such that for all $x \in \mathbb{R}$, one has*

$$\lim_{x\to\infty} \frac{d^2 f(x + ut(x))/dx^2}{d^2 f(x)/dx^2} = \exp(-u). \tag{7.288}$$

Furthermore, assume that there is also an independent integer-valued random variable for the number of losses, N satisfying $\mathbb{E}[(1+\varepsilon)^N] < \infty$, for an $\varepsilon > 0$. Then the third-order asymptotic approximation of the compound process tail probabilities, as $x \to \infty$, is given by

$$\begin{aligned} \Pr[X_1 + X_2 + \cdots + X_N \geq x] &= \overline{F}_Z(x) \\ &= a_3(x) + O(f'(x)) \end{aligned} \tag{7.289}$$

with

$$a_3(x) \ = a_2(x) - \frac{\mathbb{E}[N(X_1 + \cdots + X_{N-1})^2]}{2} f'(x). \tag{7.290}$$

Corollary 7.1 *One can show that $a_2(x)$ in Equation 7.290 is a second-order asymptotic if one makes the following two further restrictions. Consider severity distributions F satisfying the condition that*

$$\lim_{x\to\infty} \int_0^x \frac{\overline{F}(x-u)}{\overline{F}(x)} \overline{F}(u)\, du = 2\mathbb{E}[X_1] < \infty, \tag{7.291}$$

and for all $u > 0$, one considers severity density and distribution functions satisfying the following three conditions:

$$\begin{aligned} &0 < \liminf_{x\to\infty} \frac{f(ux)\overline{F}(u)}{f(u)\overline{F}(xu)} \leq \limsup_{x\to\infty} \frac{f(xu)\overline{F}(x)}{f(x)\overline{F}(xu)} < \infty, \\ &\lim_{x\to\infty} \frac{f(x+u)\overline{F}(x)}{f(x)\overline{F}(x+u)} = 1, \text{ and } \frac{\overline{F}(x/2)^2}{f(x)} = o(1). \end{aligned} \tag{7.292}$$

Now one can start to consider higher order asymptotic SLAs of the tail of the compound process distribution. These can be derived via numerous approaches, and next we consider a brief descriptions of some of these approaches as outlined in the summaries in Albrecher *et al.* (2006). We begin by presenting in Proposition 7.36 the results of Borovkov and Borovkov

(2010) which, for a given number of losses $n \in \mathbb{J}$, makes assumptions related to the properties of the severity distribution, in particular that it can be represented according to a slowly varying function. It is useful to recall the Representation Theorem in Theorem 7.28, see, for example, Bingham *et al.* (1989, theorem 1.3.1).

Theorem 7.28 (Representation Theorem for Slowly Varying Functions) *A function l is slowly varying if and only if, for some constant $a > 0$, it can be written in the form*

$$l(x) = c(x) \exp\left(\int_a^x \varepsilon(u) \frac{du}{u} \right), \; x \geq a, \tag{7.293}$$

for a measurable function $c(\cdot)$ with $c(x) \to c \in (0, \infty)$ and $\varepsilon(x) \to 0$ as $x \to \infty$.

Now if one makes the assumption that the severity distribution model can be represented according to a slowly varying function, a higher order tail asymptotic approximation can be obtained.

Proposition 7.36 (Higher Order Single-Loss Approximation via Slow Variation) *Consider a single-loss process with i.i.d losses $X_1, X_2, \ldots$ in which the severity distribution F has the properties that integer moments up to order $a \in [2, \alpha]$ are finite, $\mathbb{E}[X_1] = 0$, $\mathbb{E}[X_2^2] = 1$ and $\mathbb{E}[|X_1|^a] < \infty$, and the distribution may be represented by*

$$F(x) = x^{-\alpha} l(x) \tag{7.294}$$

with $l(x)$ a slowly varying function for some α. Furthermore, assume that given a $k \leq a$, there exist a set of slowly varying functions $\{l_j(x)\}_{0 \leq j \leq k}$ and a function $\delta_k(u, x)$ which satisfies the limit that $\lim_{(u,x) \to (0,\infty)} \delta_k(u, x) = 0$. Given such functions exist and satisfy the following decomposition of the representation of the severity distribution function

$$l(x(1+u)) = l(x) \left(\sum_{j=0}^{k-1} u^j l_j(x) + y^k l_k(x) \left(1 + \delta_k(u, x)\right) \right), \tag{7.295}$$

then the following higher order approximations of the compound distribution, given n, holds:

$$\begin{aligned} &\mathbb{P}\mathrm{r}\left[X_1 + X_2 + \cdots + X_n \geq x\right] \\ &= n\overline{F}(x) \left(1 + \sum_{j=2}^{k} \frac{(-1)^j}{x^j} U_j(x) \mathbb{E}\left[\left(\sum_{r=1}^{n-1} X_r \right)^j \right] + O\left(n^{k/2} x^{-k}\right) \right) \end{aligned} \tag{7.296}$$

with

$$U_j(x) = \sum_{m=0}^{j} \binom{-\alpha}{m} l_{j-m}(x). \tag{7.297}$$

It is clear that to generalize this result to a random number of losses now follows using, for example, the Poisson weighted mixture representations. The second approach adopted in the literature for obtaining higher order approximations of the asymptotic tail of the compound process SLA is found in works by Barbe and McCormick (2009a) and Barbe and McCormick

(2005a). This approach utilizes the concept of a smoothly varying function, see Albrecher *et al.* (2006, definition 2.1).

Proposition 7.37 (Higher Order Single-Loss Approximation via Smooth Variation) *Given a sequence of i.i.d. losses X_i with severity distribution function F and an independent integer-valued random variable for the number of losses, N. Further, under one of the following two assumptions:*

1. *the tail of the severity distribution is smoothly varying $\overline{F} \in SR_{-\rho,u}$ with index $-\rho$ and order u, then one can consider orders of approximation for any $k < \min\{\rho + 1, u + 1\}$; or*
2. *alternatively, the hazard rate of the severity distribution $h(x) = f(x)/\overline{F}(x) \in SR_{-\rho,m}$ with $\rho \in (0, 1]$ and $\lim_{x\to\infty} xh(x) = \infty$ and $\liminf_{x\to\infty} \frac{xh(x)}{\ln(x)} > 0$, then one can consider orders of approximation for any $k < m + 1$.*

Under either of these assumptions, the k-th order approximation of the annual loss distribution tails is given by

$$\Pr[X_1 + X_2 + \cdots + X_N \geq x] = a_k(x) + o\left(\frac{d^{k-1}}{dx^{k-1}}\overline{F}(x)\right), \tag{7.298}$$

with

$$a_k(x) = a_1(x) + \sum_{j=0}^{k-2} \frac{(-1)^j \mathbb{E}[N(X_1 + \cdots + X_{N-1})^{j+1}]}{(j = 1)!} \frac{d^j}{dx^j} f(x). \tag{7.299}$$

CHAPTER EIGHT

Single Loss Closed-Form Approximations of Risk Measures

This chapter is primarily focussed on the development of closed-form expressions for operational risk (OpRisk) capital and risk measures for loss distribution approach (LDA) models which are generically known as single-loss approximations (SLAs). This chapter naturally extends the asymptotic results for tails of an annual loss distribution developed in Chapter 7 to the quantile asymptotics required for risk measure approximations. To understand why this is of interest, we first discuss the nature of banking and capital accords and the changes that have been occurring in these accords over recent years which motivate one to consider developing such risk-measure-model-based approximation expressions and to study their properties. A more detailed discussion on the regulatory banking accords, their development and context was provided in companion book Cruz et al. (2015, chapters 2, 3 and 4). Before proceeding with this introduction and motivation for SLA development, we state as a summary below a few of the more important key results discussed in detail in this chapter, such as the motivation, definitions and details of these approximations are expanded upon throughout this chapter, and we simply state the results here for convenience to the reader.

8.1 Summary of Chapter Key Results on Single-Loss Risk Measure Approximation (SLA)

The following asymptotic expansions will be obtained as estimates of common risk measures, see Artzner *et al.* (1999) and McNeil *et al.* (2005). First, we present a summary of the First-order approximation results for the value-at-risk (VaR), expected shortfall (ES) and spectral risk measure (SRM):

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

1. *Value-at-Risk (VaR).* For a level $\alpha \in (0,1)$, it is given by the quantile of the annual loss distribution,

$$\begin{aligned}\mathrm{VaR}_\alpha[Z] &= F_\alpha^{\leftarrow}[Z] = \inf\{z \in \mathbb{R} : F_Z(z) \geq \alpha\} \\ &\approx F_Z^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left[1 + o(1)\right]\right) \sim F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right), \end{aligned} \tag{8.1}$$

where $F^{\leftarrow}(\bullet)$ is the generalized inverse, see Embrechts and Hofert (2010).

2. *Expected Shortfall (ES).* For a level $\alpha \in (0,1)$, the ES is given by

$$\mathrm{ES}_\alpha[Z] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s[Z]\,ds,$$

which is just the tail conditional expectation (TCE) of the annual loss distribution if distribution has no jump at $\mathrm{VaR}_\alpha[Z]$:

$$\begin{aligned}\mathrm{ES}_\alpha[Z] &= \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s[Z]\,ds = \mathbb{E}[Z|Z \geq \mathrm{VaR}_\alpha[Z]] \\ &\sim \frac{\alpha}{\alpha-1}\mathrm{VaR}_\alpha[Z] \\ &\sim \frac{\alpha}{\alpha-1}F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right), \end{aligned} \tag{8.2}$$

see Biagini and Ulmer (2009).

3. *Spectral Risk Measure (SRM).* Consider a weight function $\phi_\kappa : [0,1] \mapsto \mathbb{R}$ for some quantile level $\kappa \in [0,1]$ which decides where the weighting applies. Consider first the case with ϕ_0 where the SRM of the distribution of the annual loss random variable Z is then given by

$$\begin{aligned}\mathrm{SRM}_{\phi_0}[Z] &= \int_0^1 \phi_0(s)\mathrm{VaR}_s[Z]\,ds \\ &\sim \mathscr{K}(\alpha,\phi_1)\mathrm{VaR}_\alpha[Z] \\ &\sim \mathscr{K}(\alpha,\phi_1)F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right), \end{aligned} \tag{8.3}$$

with $\forall t \in (1,\infty)$ a new bounded weight function denoted by ϕ_1 and satisfying the bound $\phi_1(1-1/t) \leq Kt^{-1/\beta+1-\varepsilon}$ for some $K > 0$, $\beta > 0$ and $\varepsilon > 0$ where

$$\mathscr{K}(\alpha,\phi_1) = \int_1^\infty s^{1/\beta-2}\phi_1(1-1/s)\,ds. \tag{8.4}$$

For the SRM, in Tong and Wu (2012), it is shown that if an individual has a constant absolute risk aversion (CARA) utility function with coefficient of absolute risk aversion ξ, then the SRM with respect to some quantile level $\kappa \in [0,1]$ should be given as

$$SRM_{\phi_\kappa}[Z] = \int_\kappa^1 \phi_\kappa(s)\mathrm{VaR}_s[Z]ds,$$

where

$$\phi_\kappa(s) := (1-\kappa)^{-1}\phi_1\left(1 - \frac{1-s}{1-\kappa}\right)I_{[\kappa,1]}(s),$$

with not just a bound on the function ϕ_1 but an explicit expression in this case given by

$$\phi_1(\kappa) = \frac{\xi e^{-\xi(1-\kappa)}}{1 - e^{-\xi}}.$$

A second case that arises often in practice where one can assume knowledge of the function ϕ_1 is the case that recovers the risk measure corresponding to the ES. Note that if one takes $\phi_1(t) \equiv 1, \forall t \in [0, 1]$, then the SRM reduces to the ES.

A summary of the second-order asymptotic expansion results discussed in this chapter is provided next for the VaR, ES and SRM. In the case that one assumes that severity distribution F_X has finite mean and the hazard rate $h_X(x) = \frac{f_X(x)}{1-F_X(x)}$ is of regular variation $h_X \in RV_{-\beta}$ for $\beta \geq 0$, then as $\alpha \to 1$, one has for the inverse of the annual loss distribution the result (see Albrecher *et al.* 2010 and Degen 2010)

$$\mathrm{VaR}_\alpha[Z] = F_Z^{-1}(\alpha) = F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left\{1 + \widetilde{c}_\beta g_1\left(F_X^{-1}(\widetilde{\alpha})\right) + o(g_1(F_X^{-1}(\widetilde{\alpha})))\right\}^{-1}\right), \tag{8.5}$$

where $\widetilde{\alpha} = 1 - (1-\alpha)/\mathbb{E}[N]$ and

$$g_1(x) = \begin{cases} \dfrac{f_X(x)}{1 - F_X(x)}, & \text{if } \mathbb{E}[X] < \infty, \\ \dfrac{\int_0^x \overline{F}_X(s)ds f_X(x)}{1 - F_X(x)}, & \text{if } \mathbb{E}[X] = \infty, \end{cases}$$

$$\widetilde{c}_\beta = \begin{cases} \dfrac{\mathbb{E}[X]\mathbb{E}[(N-1)N]}{\mathbb{E}[N]}, & \text{if } \mathbb{E}[N] < \infty, \\ \dfrac{c_\beta \mathbb{E}[(N-1)N]}{\mathbb{E}[N]}, & \text{if } \mathbb{E}[N] = \infty, \end{cases}$$

where we define

$$c_\beta = \begin{cases} 1, & \text{if } \beta = 1, \\ (1-\beta)\frac{\Gamma^2\left(1-\frac{1}{\beta}\right)}{2\Gamma\left(1-\frac{2}{\beta}\right)}, & \text{if } \beta \in (1, \infty). \end{cases} \tag{8.6}$$

For the second-order result for the SRM for a given tail function $\overline{F}_X(t) \in RV_{-\alpha}$, define $U_F(t) := (1/(1-F))^{\leftarrow}(t) \in RV_{1/\alpha}$. Now assume that for some $\rho \leq 0, \exists A(t) \in RV_\rho$ such that

$$\lim_{t \to +\infty} \frac{1}{A(t)}\left(\frac{U_F(tx)}{U_F(t)} - x^{1/\alpha}\right) = x^{1/\alpha}\frac{x^\rho - 1}{\rho}.$$

Then,

$$SRM_{\phi_\kappa}[Z] \sim \left[\mathscr{K}(\alpha, \phi_1) + A\left(\frac{\mathbb{E}[N]}{1-\kappa}\right)\mathscr{M}(\alpha, \phi_1, \rho)\right] F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right),$$

as $\kappa \to 1$, where

$$\mathscr{M}(\alpha, \phi_1, \rho) = \frac{1}{\rho}\int_1^{+\infty} t^{1/\alpha - 2}(t^\rho - 1)\phi_1\left(1 - \frac{1}{t}\right) dt$$

and $\mathscr{K}(\alpha, \phi_1)$ as defined in Equation 8.4, see details in Tong and Wu (2012).

This result then naturally admits the second-order result for the ES asymptotic expansion using the setting of constant utility noted above in which $\phi_1(t) = 1, \forall t \in [0,1]$. Then, when $\kappa \to 1$, we have the following second-order approximation of ES_κ, see Tong and Wu (2012):

$$\mathrm{ES}_\kappa \sim \left[\frac{\alpha}{\alpha-1} + \frac{\alpha^2}{(\alpha-\alpha\rho-1)(\alpha-1)} A\left(\frac{\mathbb{E}[N]}{1-\kappa}\right)\right] F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right).$$

The properties of such asymptotic SLA estimates are still an active subject of study with regard to aspects such as explicit approximation errors, unbiased quantile function estimation, asymptotic rates of convergence, sensitivity to parameter estimation and model misspecification. In the remainder of this chapter, we detail these asymptotic expansions and their properties as well as providing examples of their approximation performance. We start with motivation for development of such approximations. To motivate the direct utility of such risk measure asymptotic approximations we first present a discussion on how the Basel II/III banking accords have developed and focused on aspects of such quantitative measures of risk to be considered for capital adequacy of banks.

8.2 Development of Capital Accords and the Motivation for SLAs

In jurisdictions in which active regulation is applied to the banking sector, the modeling of OpRisk has progressively taken a prominent place in financial quantitative measurement. This has occurred as a result of Basel II and now Basel III regulatory requirements. There has been a significant amount of research dedicated to understanding the features of Basel II, see, for example, Daniélsson *et al.* (2001), Decamps *et al.* (2004) and Kashyap and Stein (2004). In addition, the mathematical and statistical properties of the key risk processes that comprise OpRisk, especially those that contribute significantly to the capital charge required to be held against OpRisk losses, have also been carefully studied, see, for example, the book-length discussions in Cruz (2002), King (2001) and Shevchenko (2011).

This chapter will provide a clear and concise understanding of several increasingly popular approaches to capital approximation from the perspective of developing analytic closed-form expressions. In the process, we demonstrate several important key details of such capital approximations, discuss the implications of use of such approximations and their shortcomings and assess their behaviours in simple but realistic loss models adopted in practice.

In particular, we provide a detailed discussion on the attributes of the so-called capital approximation methods known as single-loss-type approaches of first- and second-order expansions. In doing so, this chapter draws together several areas of the literature to allow industry professionals an insight into the processes that have lead naturally from the results on tail asymptotics in Chapter 7 to expansions of tail functionals for risk measures such as VaR, ES and SRMs considered in this chapter. This will be undertaken in the context of light- and heavy-tailed loss process models, in similar manner to the sections developed in Chapter 7 for tail asymptotic approximations.

To appreciate why such focus has been placed on developing an understanding of such asymptotics, these results are required to be placed in context. To motivate these developments from a practical perspective, we discuss first the regulatory evolution of OpRisk modeling frameworks. In January 2001, the Basel Committee on Banking Supervision (BCBS) proposed the Basel II Accord (BCBS, 2002; BCBS, 2004; BCBS, 2006) which replaced the 1988 Capital Accord. Now, in 2013, the Basel III Accord is due to start to be considered. Since the initiation

of the Basel capital accords, the discipline of OpRisk and its quantification have grown in prominence in the financial sector. Paralleling these developments have been similar regulatory requirements for the insurance industry which are referred to as Solvency II. In both accords, a primary component of such regulation revolves around the quantitative modeling of capital.

Under the Basel II/III structures, there is at the core the notion of three pillars, which, by their very nature, emphasize the importance of assessing, modeling and understanding OpRisk loss profiles. These three pillars are minimum capital requirements (refining and enhancing risk modeling frameworks), supervisory review of an institution's capital adequacy and internal assessment processes and market discipline, which deals with disclosure of information.

In the third update to the Basel Accords due for implementation in the period 2013–2018, a global regulatory standard which draws together bank capital adequacy, stress testing and market liquidity is created. It is established as an international best practice for modeling OpRisk by the members of the BCBS, see Gregoriou (2009) and discussions in Blundell-Wignall and Atkinson (2010).

The Basel III accord naturally extends on the work developed in both the Basel I and Basel II accords with the new accord arising primarily as a response to the identified issues associated with financial regulation that arose during the recent global financial crisis in the late-2000s. In this regard, the Basel III accord builds on Basel II by strengthening the bank capital requirements as well as introducing additional regulatory requirements on bank liquidity and leverage.

Banking regulation under Basel II and Basel III specifies that banks are required to hold adequate capital against OpRisk losses. OpRisk is a relatively new category of risk which is additional to more well-established risk areas such as market and credit risks. As such in its own right OpRisk attracts a capital charge which is defined by Basel II/III (BCBS 2006, p.144) as

"*[. . .] the risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. This definition includes legal risk, but excludes strategic and reputational risk.*"

OpRisk is significant in many financial institutions. Although this chapter is focused primarily on the quantification and approximation of OpRisk capital measures, it is important to realise that such capital measures must then be realised in the form of assets held by the financial institution of bank. In this section we briefly discuss the different classifications of assets that must be performed under the Basel II/III standards as well as discussing how these have evolved from the Basel I to the Basel III frameworks. Before detailing the changes to capital requirements due to come into industry practice under Basel III, it is prudent to recall the Basel definition of Tier I capital, which is the key measure of a bank's financial strength from the perspective of the regulatory authority. In particular, the capital accord in Basel II and III states that financial institutions must provide capital above the minimum required amount, known as the floor capital. In addition, this capital as specified in regulation is comprised three key components, Tier I, Tier II and Tier III. Both Tier I and Tier II capital were first defined in the Basel I capital accord and remained substantially the same in the replacement Basel II and Basel III accords.

Definition 8.1 (Tier I Capital) *The Tier I capital under regulation is comprised the following main components:*

1. *Paid-up share capital/common stock;*
2. *Disclosed reserves (or retained earnings).*

It may also include non-redeemable non-cumulative preferred stock.

The Basel Committee also noted the existence of banking strategies to develop instruments in order to generate Tier I capital. As a consequence, these must be carefully regulated through imposition of stringent conditions, with a limit to such instruments at a maximum of 15% of total Tier I capital.

Definition 8.2 (Tier II Capital) *The Tier II capital under regulation is comprised the following main components:*

1. *Undisclosed reserves;*
2. *Asset revaluation reserves;*
3. *General provisions/general loan-loss reserves;*
4. *Hybrid (debt/equity) capital instruments; and*
5. *Long-term subordinated debt.*

In this regard, one may consider Tier II capital as representing the so-called 'supplementary capital'.

We note at this stage that as a consequence of different legal systems in each jurisdiction, the accord has had to be sufficiently flexible to allow for some interpretation of specific capital components within the context of each regulators jurisdiction. Depending on the particular jurisdiction in question, the specific country's banking regulator has some discretionary control over how exactly differing financial instruments may count in capital calculations.

Remark 8.1 *The key reason that Basel III requires financial institutions to hold capital is that it is aimed to provide protection against unexpected losses. This is different to mitigation of expected losses, which are covered by provisions, reserves and current year profits.*

We note that modifications under the Basel III accord relative to its predecessor refer to limitations on risk-weighted capital (RWC) and the Tier I capital ratio, as defined below.

Definition 8.3 (Risk-Weighted Assets) *These assets comprise the total of all assets held by the bank weighted by credit risk according to a formula determined by either the jurisdictions regulatory authority or, in some cases, the central bank. Most regulators and central banks adhere to the definitions specified by the BCBS guidelines in setting formulae for asset risk weights. Liquid assets such as cash and coins typically have zero risk weight, whilst certain loans have a risk weight at 100% of their face value. As specified by the BCBS, the total RWA is not limited to credit risk. It contains components for market risk (typically based on VaR) and OpRisk. The BCBS rules for calculation of the components of total RWA have also been updated as a result of the recent financial crisis.*

Definition 8.4 (Tier I Capital Ratio) *The Tier 1 capital ratio is the ratio of a bank's core equity capital to its total risk-weighted assets (RWA).*

Next we highlight the prominent extensions to the Basel II accord, established in the Basel III accord. In particular, the Basel III accord will require financial institutions to hold for risk-weighted assets 4.5% of common equity which is an increase from the previous 2% under Basel II as well as 6% of Tier I capital itself an increase by 2% relative to Basel II. In addition to these changes to common equity and Tier I capital, Basel III also introduces a minimum leverage ratio and two additional required liquidity ratio limits. Finally, of the significant changes, there are also additional capital buffers introduced:

1. Introduction of a mandatory capital conservation buffer of 2.5% ; and

2. A discretionary countercyclical buffer, allowing national regulators to require up to another 2.5% of capital during periods of high credit growth.

Against the backdrop of these capital regulatory accord changes and extensions, there is always the base fundamental requirement of risk analysts, actuaries and quants which involves the quantitative modeling and reporting of such capital estimates. To quantify the OpRisk capital charge under the current regulatory framework for banking supervision, referred to as Basel II/Basel III, many banks adopt the LDA. In this context, we are working with compound processes and partial sums as considered for tail asymptotic analysis in Chapter 7. In this chapter, the primary concern involves the development of closed-form approximation and estimation of OpRisk capital and risk measures for LDA models. There are typically three main families of risk measure that are considered for the calculation of OpRisk capital: VaR (LDA annual loss distribution quantile function); ES (LDA annual loss distribution tail conditional expectation (TCE)) and SRM. Note that there is a detailed discussion on OpRisk risk measures in Cruz *et al.* (2015, chapter 6); here we simply reiterate the definition of these three choices for clarity in future developments of this chapter, see discussions in Artzner *et al.* (1999).

Definition 8.5 (Risk Measures: VaR, ES and Spectral Risk Measures) *Consider an LDA model for an OpRisk single-loss process with annual loss random variable* $Z_N = X_1 + \cdots + X_N \sim F_{Z_N}(z)$ *with severity distribution* $X_i \sim F_X(x)$ *for all losses* X_i *and frequency distribution* $N \sim F_N(n)$. *The risk measures for VaR, ES and SRM are then defined by*

Value-at-Risk (VaR) *is defined for an LDA OpRisk annual loss random variable* Z_N, *for a level* $\alpha \in (0,1)$, *via the quantile of the annual loss distribution,*

$$\mathrm{VaR}_\alpha[Z_N] = F^{\leftarrow}_{Z_N}(\alpha) = \inf\{z \in \mathbb{R} : F_{Z_N}(z) \geq \alpha\}, \tag{8.7}$$

where $F^{\leftarrow}(\cdot)$ *is the generalized inverse, see Embrechts and Hofert (2010).*

Expected Shortfall (ES) *for a level* $\alpha \in (0,1)$ *is given by the tail expectation of the annual loss distribution according to*

$$\mathrm{ES}_\alpha[Z_N] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s[Z_N]ds = \mathbb{E}[Z_N | Z_N \geq \mathrm{VaR}_\alpha[Z_N]], \tag{8.8}$$

see Biagini and Ulmer (2009). The second equality is valid if distribution has no jump at $\mathrm{VaR}_\alpha[Z_N]$.

Spectral Risk Measure (SRM) *for a weight function* $\phi : [0,1] \mapsto \mathbb{R}$ *is given by*

$$\mathrm{SRM}_\phi[Z] = \int_0^1 \phi(s)\mathrm{VaR}_s[Z_N]ds. \tag{8.9}$$

Remark 8.2 *For the SRM, in Tong and Wu (2012), it is shown that if an individual has a CARA utility function with coefficient of absolute risk aversion* ξ, *then the SRM should be given for quantile level* $\kappa \in [0,1]$ *as*

$$SRM_{\phi_\kappa}[Z] = \int_\kappa^1 \phi_\kappa(s)\mathrm{VaR}_s[Z_N]ds,$$

where

$$\phi_\kappa(s) = (1-\kappa)^{-1}\phi_1\left(1-\frac{1-s}{1-\kappa}\right)I_{[\kappa,1]}(s)$$

with $\forall t \in (1,\infty)$ *a function denoted by* ϕ_1 *that satisfies the bound* $\phi_1(1-1/t) \leq Kt^{-1/\beta+1-\varepsilon}$ *for some* $K > 0$ *and* $\varepsilon > 0$. *and in the case of the CARA utility function can be parametrically expressed according to*

$$\phi_1(\kappa) = \frac{\xi e^{-\xi(1-\kappa)}}{1-e^{-\xi}}.$$

Note that if one takes $\phi_1(t) \equiv 1$, $\forall t \in [0,1]$, *then the SRM resumes to the ES.*

Considering these three risk measures as defined earlier, this chapter extends the asymptotic results presented in Chapter 7 to also develop risk measure approximations. Hence, studies of different asymptotic approximations are considered for the estimation of risk measures and tail properties of risk processes in single risk modeling scenarios. The first section discusses the family of SLAs in their basic form as well as the recent refinements to such estimates for VaR and also for ES. This is studied in both the light-tailed severity and sub-exponential heavy-tailed severity model contexts. In comparison, the VaR estimates obtained from extreme value theory (EVT) are considered and compared. Then a discussion of the asymptotics for SRM in an OpRisk setting is developed. Summarizing, the following sections will consider single risk loss process asymptotics/SLAs:

- Closed-form quantile functions and conditional expectations for OpRisk severity and LDA models;
- Fractional series expansions for risk measure estimation;
- Closed-form approximation for quantile risk measures VaR (standard estimate);
- Approximation for quantiles VaR (refinements);
- Penultimate approximations for quantile risk measures VaR: EVT SLA asymptotics (penultimate approximations): the EVT-based asymptotic estimators for standard linear normalized extreme value domains of attraction and power-normalized extreme value domains of attraction;
- Perturbation approximations for VaR;
- Approximation for ES (standard estimate);
- Approximation for SRM;
- Quantification of approximation error, sensitivity and model structure uncertainty in SLA-VaR.

8.3 Examples of Closed-Form Quantile and Conditional Tail Expectation Functions for OpRisk Severity Models

In developing SLAs of first- and second-order compound process loss models, the first step will be to consider classes of models that admit a closed-form or series expansion expression for the quantile function of the severity model. These severity model quantile functions will then be utilized in the capital approximation for the compound process. Therefore, this section starts

by presenting some known results regarding classes of OpRisk models that were discussed in previous chapters.

We start by discussing some results developed in Landsman and Valdez (2005). In Chapter 4, the exponential dispersion family (EDF) of severity distribution models were considered; these are also discussed in the context of closed-form LDA models in Chapter 6. First, we present some known results regarding the closed-form representations of the asymptotic quantile and TCE or ES. The TCE or ES represents the average loss given exceedance of some specified tail quantile and is, therefore, a measure of the average incurred loss beyond some critical threshold that a financial institution could be exposed to.

The TCE given by Equation 8.16 is studied in numerous places in the actuarial literature, such as Panjer (2002) where the TCE is developed for the multivariate Gaussian family and the extended TCE results for the class of elliptical distributions studied in Landsman and Valdez (2003).

Remark 8.3 *The following VaR and ES closed-form results for different loss model families is of direct relevance to the first- and second-order asymptotic approximations of these risk measures for the LDA compound processes that contain these loss models as their severity. This is because the aim of the SLA representation is to obtain an expression for the compound process capital (tail functional) in terms of the corresponding tail functionals of the severity loss model for which the density, distribution or quantile function may be closed form. In other cases where the loss model used for the severity does not admit closed-form expressions, we also provide a brief section on the series expansion approximation as well as empirical process approximations.*

8.3.1 EXPONENTIAL DISPERSION FAMILY LOSS MODELS

We start this section with a brief discussion on the VaR for a special sub-family of the EDF loss models discussed in Chapter 4, the inverse Gaussian (IG) sub-family. For the generalization of the IG distribution given by the generalized inverse Gaussian (GIG) model, see discussion in Chapter 4; the quantile function and, therefore, the VaR expression for a given reporting level α were developed in an infinite series expansion in Munir and Shaw (2012). In Chapter 4, we discussed how the IG distribution was a member of the GIG model when $p = -\frac{1}{2}$. Therefore, the VaR expression in terms of a convergent infinite series is given by Proposition 8.1, see details in Munir and Shaw (2012).

Proposition 8.1 (VaR for Inverse Gaussian Sub-Family of EDF Loss Models) *Consider the IG loss model with density given by*

$$f_X(x;\mu,\lambda) = \left[\frac{\lambda}{2\pi x^3}\right]^{1/2} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0. \tag{8.10}$$

This IG loss model is then a sub-family of the GIG loss model given by

$$f_X(x;a,b,p) = \frac{\left(\frac{a}{b}\right)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} \exp\left(-\frac{ax + b/x}{2}\right) \tag{8.11}$$

when $p = -\frac{1}{2}$, $a = \lambda/\mu^2$ and $b = \lambda$. The quantile function for the GIG model is then given by a Taylor series expansion (for $\lambda \neq 1$) around a value $\alpha_q \in (0,1)$ such that $F_X^{-1}(\alpha_q) = x_q$ according to the convergent polynomial series of one of the two forms depending on the value of λ given by

1. *If $\lambda \neq 1$, then one has the series expansion:*

$$\mathrm{VaR}_{\alpha_q}[X] = \sum_{n=0}^{\infty} c_n(\alpha - \alpha_q)^n \tag{8.12}$$

with

$$c_n = \begin{cases} x_q, & n = 0, \\ \dfrac{2}{n} K_\lambda(\omega) \displaystyle\sum_{i=0}^{n-1} b_i d_{n-i-1}, & n \geq 1, \end{cases}$$
$$d_n = \begin{cases} c_0^{1-\lambda}, & n = 0, \\ \dfrac{1}{c_0} \displaystyle\sum_{i=1}^{n} \left(\dfrac{(2-\lambda)i}{n} - 1 \right) c_i d_{n-i}, & n \geq 1. \end{cases} \tag{8.13}$$

2. *If $\lambda = 1$, then one has the series expansion*

$$\mathrm{VaR}_{\alpha_q}[X] = \sum_{n=0}^{\infty} c_n (\alpha - \alpha_q)^n \tag{8.14}$$

with

$$a_n = \begin{cases} \dfrac{1}{c_0}, & n = 0, \\ -\dfrac{1}{c_0} \displaystyle\sum_{i=1}^{n} c_i a_{n-i}, & n \geq 1, \end{cases}$$
$$b_n = \begin{cases} \exp\left(\dfrac{\omega}{2} (a_0 + c_0) \right), & n = 0, \\ \dfrac{\omega}{2n} \displaystyle\sum_{i=1}^{n} i(a_i + c_i) b_{n-i}, & n \geq 1, \end{cases} \tag{8.15}$$
$$c_n = \begin{cases} x_q, & n = 0, \\ (2/n) K_1(\omega) b_{n-1}, & n \geq 1, \end{cases}$$

where the reparameterizations $\nu = \sqrt{\frac{a}{b}} = 1$ and $\omega = \sqrt{ab}$ are utilized in the expansions obtained earlier.

Expanding upon this approach for the VaR representation (quantile function) of a GIG, one could restrict $p = -\frac{1}{2}$ to get the IG sub-family and then integrate the terms in the series expansion over some restricted interval of probabilities for α to obtain the ES for this IG sub-family of the EDF. The validity of this approach could then be studied in terms of the convergence of the resulting infinite series expansion.

Alternatively, one can develop a closed-form expression for ES for the general EDF loss models as presented below. The wider family of EDF loss models can have a general expression for the ES developed, which allows all sub-families of this loss model class to be included such as the IG above. Here we focus on the TCE results developed for the EDFs in Landsman and Valdez (2005). Consider a loss random variable $X \sim F_X(x)$, then the ES or TCE can be written in the following form:

$$\mathrm{ES}_q[X] = TCE_q[X] = \frac{1}{\overline{F}_X(x_q)} \int_{x_q}^{\infty} x dF_X(x) \tag{8.16}$$

when $\overline{F}_X(x_q) > 0$ for a quantile value x_q satisfying $\overline{F}_X(x_q) = 1 - q$ with the quantile given by $x_q = \inf\{x|F_X(x) \geq q\}$ and there is no jump in distribution $F(x)$ at $x = x_q$.

Given this definition, one can then define the TCE for the EDF according to the results in Theorem 8.10.

Theorem 8.1 (Closed-Form ES for Exponential Dispersion Families) *Consider a loss random variable $X \sim F_X(x)$ where $F_X(x) = EDF(\theta, \lambda)$ is an EDF distribution with canonical parameter θ and index parameter λ. The EDF distribution is characterized by its moment-generating function given by*

$$mgf_X(s) = \exp\left(\lambda\left[\kappa\left(\theta + \frac{1}{\lambda}s\right) - \kappa(s)\right]\right) \tag{8.17}$$

with mean and variance given by

$$\begin{aligned} \mu &= \mathbb{E}[X] = \tau(\theta) = \kappa'(\theta), \\ \mathbb{V}\text{ar}(X) &= \tfrac{1}{\lambda}\kappa''(\theta). \end{aligned} \tag{8.18}$$

For an EDF distribution with a generating normal exponential family baseline distribution which satisfies that it is regular or steep, then the TCE can be expressed by the following results:

1. *For the reproductive EDF sub-family, the TCE is given by*

$$TCE_q[X] = \mu + \sigma^2 h, \tag{8.19}$$

where $\sigma = \frac{1}{\lambda}$ represents the dispersion parameter and h is given according to the derivative of the distribution tail known as the generalized hazard function given by

$$h := \frac{\partial}{\partial\theta}\ln\overline{F}_X(x_q|\theta, \lambda). \tag{8.20}$$

2. *For the additive EDF sub-family, the TCE is given by*

$$TCE_q[X] = \mu + h. \tag{8.21}$$

One can then consider a few example sub-families of this class of EDF distributions for which the TCE is known in closed form, see discussion in Landsman and Valdez (2005).

EXAMPLE 8.1 Inverse Gaussian EDF Sub-Family Closed-Form TCE

Consider a loss random variable X from an IG sub-family of the EDF class given by $X \sim IG(\lambda, \mu)$ with density and distribution

$$f_X(x; \lambda, \mu) = \sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0, \tag{8.22}$$

$$F_X(x; \lambda, \mu) = \Phi\left(\frac{1}{\mu}\sqrt{\lambda x} - \sqrt{\frac{\lambda}{x}}\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\frac{1}{\mu}\sqrt{\lambda x} - \sqrt{\frac{\lambda}{x}}\right)$$

with the cumulant function

$$\kappa(\theta) = -\frac{1}{\mu} = -\sqrt{(-2\theta)}, \tag{8.23}$$

where $\theta = -\frac{1}{2\mu^2}$. The TCE for the loss random variable X is then given by

$$\mathrm{TCE}_X(x_q) = \mu + \frac{\mu}{\lambda \overline{F}(x_q|\mu,\lambda)}\sqrt{\lambda x_q}\phi(z_q^*) + \frac{\mu}{\lambda \overline{F}(x_q|\mu,\lambda)} \times \left\{ \exp\left(\frac{2\lambda}{\mu}\right)\left[2\lambda\Phi\left(-z_q^* - 2\sqrt{\frac{\lambda}{x_q}}\right) - \sqrt{\lambda x_q}\phi\left(-z_q^* - 2\sqrt{\frac{\lambda}{x_q}}\right)\right]\right\}, \tag{8.24}$$

where

$$z_q^* = \frac{1}{\mu}\sqrt{\lambda x_q} - \sqrt{\frac{\lambda}{x_q}} \tag{8.25}$$

and $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal distribution and its density. ■

Remark 8.4 *The results just presented correspond to the tail conditional expectation approximations for the single severity distribution in the EDF class. However, in practice we are interested in the case of partial sums and random sums. However, as will be presented below, in special cases such as the EDF family, the single distributional results developed above are sufficient to make the required extension to the partial sums.*

In the case of partial sums of i.i.d. loss random variables with a common severity distribution given by the EDF, one can still derive the TCE for this family of models. In this case, one obtains the result given in Proposition 8.2.

Proposition 8.2 (Closed-Form ES for Exponential Dispersion Family Partial Sums) *Consider n i.i.d. losses $X_1, X_2, \ldots, X_n$ with each loss given by the same distribution $X_i \sim F_X(x;\theta_i,\lambda_i)$ given by an additive exponential family form with means μ_i, then the TCE of the partial sum*

$$Z_n = \sum_{i=1}^{n} X_i \sim F_{Z_n}(z), \tag{8.26}$$

is given by

$$\mathrm{TCE}_q[Z_n] = \mathrm{TCE}_q[X_i; z_q, \theta_n, \lambda_n], \tag{8.27}$$

with z_q the q quantile of the distribution of the partial sum which has parameters

$$\mu_n = \sum_{i=1}^{n}\mu_i, \quad \lambda_n = \sum_{i=1}^{n}\lambda_i. \tag{8.28}$$

Remark 8.5 *Note that in Chapter 4, it was discussed that in some sub-families of the EDF, the distribution for the compound process LDA structure could be obtained in closed form; in these cases, the above results for the VaR and ES will also coincide with closed-form expressions for the LDA model compound process risk measures. In such special cases, the first- and second-order SLA representations developed below will not be required.*

Next we will consider the families of loss models based on Tukey transform loss models and their generalizations, as discussed in Chapter 4. In particular, we consider the g-and-h

distributions and the closed-form expressions that are known for the VaR and ES for such loss models that will be of relevance for the expression of the SLA for a compound process comprised this severity model in this family.

8.3.2 *g*-AND-*h* DISTRIBUTION FAMILY LOSS MODELS

In Chapter 4, the g-and-h family of distributions were described; here we recall some known results regarding the closed-form expressions these loss models admit for the quantile function (VaR), see details in Nam and Gup (2003) and Jiménez and Arunachalam (2011). Recall, in Chapter 4, the g-and-h family of loss models were defined according to the transformation reiterated in Definition 8.6.

Definition 8.6 (*g*-and-*h* Loss Models) *Consider the loss model for the g-and-h family defined by the transformation of a standard Gaussian random variable* $Y \sim Normal(0,1)$ *according to*

$$X = T_{g,h}(Y) = \frac{1}{g}(\exp(gY) - 1)\exp\left(\frac{1}{2}hY^2\right), \quad g \neq 0, \quad h \in \mathbb{R}, \tag{8.29}$$

where the parameters g *and* h *are the bias and elongation. One may also scale and translate* X *by translation parameter* A *and scale parameter* $B > 0$ *to obtain a new g-and-h loss distribution given by*

$$X = T_{g,h}(Y; A, B) = A + BT_{g,h}(Y). \tag{8.30}$$

The quantile function (VaR) for the g-and-h loss model was discussed in Chapter 4 and is trivial to obtain because of the representational form of this loss model. We reiterate the result in Proposition 8.6.

Proposition 8.3 (Closed-Form VaR for *g*-and-*h* Loss Models) *Consider a loss model characterized by a g-and-h distribution* $X \sim F_X(x; g, h, A, B)$. *The VaR for some level* $\alpha \in (0,1)$ *is given in closed form by*

$$\begin{aligned} \text{VaR}_\alpha[X] &= A + BT_{g,h}(z_\alpha), \\ \text{VaR}_{1-\alpha}[X] &= A - B\exp(-gz_\alpha)T_{g,h}(z_\alpha), \end{aligned} \tag{8.31}$$

for the α*-th quantile of the standard Gaussian given by* $z_\alpha = \Phi^{-1}(\alpha)$.

In order to obtain a closed-form result for the ES of the g-and-h loss distributions, it will be useful to recall the following result in Proposition 8.4, see details in Jiménez and Arunachalam (2011).

Proposition 8.4 *Consider a continuous-valued loss random variable* X *with distribution* $X \sim F_X(x)$ *and density* $f_X(x)$. *If the derivative of the distribution is strictly non-zero, then the inverse distribution is differentiable and satisfies*

$$\int_\alpha^1 F_X^{-1}(u)du = \int_{x_\alpha}^\infty w f_X(w)dw, \tag{8.32}$$

where u *satisfies* $F_X(x_u) = u$.

One can then utilize this result to derive closed-form expressions for the ES under the g-and-h loss model according to the results in Proposition 8.5, see Jiménez and Arunachalam (2011).

Proposition 8.5 (Closed-Form ES for *g*-and-*h* Loss Models) *Consider a loss model characterized by a g-and-h distribution $X \sim F_X(x; g, h, A, B)$. The ES for some level $\alpha \in (0,1)$ is given in closed form by*

$$\begin{aligned} \mathrm{ES}_\alpha[X] = A + \frac{B}{1-\alpha} \frac{\exp\left(\frac{g^2}{2(1-h)}\right)}{g\sqrt{1-h}} \left[1 - \Phi\left(\sqrt{1-h} z_\alpha - \frac{g}{\sqrt{1-h}}\right)\right] \\ - \frac{B}{1-\alpha}\left[\frac{1}{g\sqrt{1-h}}\left(1 - \Phi\left(\sqrt{1-h} z_\alpha\right)\right)\right], \end{aligned} \tag{8.33}$$

with $\Phi(\cdot)$ the standard normal distribution and z_α the α-th quantile of the standard normal distribution.

Similar results are then obtained for the sub-families of g, h and generalized g-and-h distributions. Next we consider a generalization of the generalized Pareto distribution (GPD) heavy-tailed model, discussed in Chapter 4, which involves considering the family of asymmetric power distribution (APD) loss models.

8.3.3 EXTENDED GPD: THE ASYMMETRIC POWER FAMILY LOSS MODELS

Komunjer (2007) considered a family of distributions given by the APD which is a generalization of the GPD family with parameters: λ and asymmetry parameter $\alpha \in (0,1)$. Note that the GPD family arises when $\alpha = \frac{1}{2}$ and the general definition for the APD family of loss models is given in Definition 8.7.

Definition 8.7 (Asymmetric Power Distribution Loss Models) *A loss random variable X with density $f_X(x; \lambda, \alpha)$ is distributed according to a standard APD family if the density is given by*

$$f_X(x;\lambda,\alpha) = \begin{cases} \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left(-\dfrac{\delta_{\alpha,\lambda}}{\alpha^\lambda}|x|^\lambda\right), & \text{if } x \le 0, \\ \dfrac{\delta_{\alpha,\lambda}^{1/\lambda}}{\Gamma(1+1/\lambda)} \exp\left(-\dfrac{\delta_{\alpha,\lambda}}{(1-\alpha)^\lambda}|x|^\lambda\right), & \text{if } x > 0, \end{cases} \tag{8.34}$$

with $\alpha \in (0,1)$, $\lambda > 0$ and

$$\delta_{\alpha,\lambda} = \frac{2\alpha^\lambda(1-\alpha)^\lambda}{\alpha^\lambda + (1-\alpha)^\lambda}. \tag{8.35}$$

The distribution function is given by

$$F_X(x) = \begin{cases} \alpha\left[1 - I\left(\delta_{\alpha,\lambda}\alpha^{-\lambda}\sqrt{\lambda}|x|^\lambda, 1/\lambda\right)\right], & \text{if } x \le 0, \\ 1 - (1-\alpha)\left[1 - I\left(\delta_{\alpha,\lambda}\alpha^{-\lambda}\sqrt{\lambda}|x|^\lambda, 1/\lambda\right)\right], & \text{if } x > 0, \end{cases} \tag{8.36}$$

and Pearson's (1922) incomplete gamma function given by

$$I(x,y) = \frac{1}{\Gamma(Y)} \int_0^{x\sqrt{y}} t_{y-1} \exp(-t) dt. \tag{8.37}$$

The mean and variance of the standard APD loss model are given by

$$\begin{aligned} \mathbb{E}[X] &= \frac{\Gamma\left(\frac{2}{\lambda}\right)}{\Gamma\left(\frac{1}{\lambda}\right)}(1-2\alpha)\delta_{\alpha,\lambda}^{-\frac{1}{\lambda}}, \\ \mathbb{V}\text{ar}[X] &= \frac{\Gamma\left(\frac{3}{\lambda}\right)\Gamma\left(\frac{1}{\lambda}\right)(1-3\alpha+3\alpha^2) - \Gamma\left(\frac{2}{\lambda}\right)^2(1-2\alpha)^2}{\left[\Gamma\left(\frac{1}{\lambda}\right)\right]^2}\delta_{\alpha,\lambda}^{-\frac{2}{\lambda}}. \end{aligned} \tag{8.38}$$

The standard APD loss distribution can be transformed by affine transforms to obtain a linearly scaled and translated APD distribution $\widetilde{X} \sim F_{\widetilde{X}}(x;\lambda,\alpha,\theta,\phi)$ *with* $\widetilde{X} = \theta + \phi X$.

Although this distribution takes negative support, it can be truncated or utilized in splice models as discussed in Chapter 1 in which case the ES and VaR closed-form expressions presented below may be modified for the truncation. Alternatively, the asymmetry parameter may be selected to minimize the probability attributed by the model to a negative loss, making such tail functionals still highly accurate. Note that the ES will always be accurate for truncations, just the loss value (quantile value) at which the integration domain starts would be adjusted for truncation.

Under this family of distribution, the VaR and ES can be found in closed form. The quantile function, which will yield the VaR, is given in Proposition 8.6 and the ES or TCE is given in Proposition 8.7

Proposition 8.6 (Quantile Function for APD Loss Model: VaR) *A loss random variable* X *with density* $f_X(x;\lambda,\alpha)$ *given by a standard APD family has a quantile function given for any probability level* $u \in (0,1)$ *by the inverse distribution*

$$F_X^{-1}(u;\lambda,\alpha) = \begin{cases} -\left[\dfrac{\alpha^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\dfrac{u}{\alpha},\dfrac{1}{\lambda}\right)\right]^{1/\lambda}, & \text{if } u \le \alpha, \\ \left[\dfrac{(1-\alpha)^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\dfrac{1-u}{1-\alpha},\dfrac{1}{\lambda}\right)\right]^{1/\lambda}, & \text{if } u > \alpha, \end{cases} \tag{8.39}$$

where $I^{-1}(x,y)$ *is the inverse of Pearson's 1922 incomplete gamma function given by solving* $x = I^{-1}(s,y)$ *which is equivalent to* $s = I(x,y)$.

The TCE for the APD family is given by Proposition 8.7, see details in Komunjer (2007).

Proposition 8.7 (Tail Conditional Expectation for APD Loss Model: ES) *A loss random variable* X *with density* $f_X(x;\lambda,\alpha)$ *given by a standard APD family has a TCE (or ES) given for any probability level* u *by one of the following two forms:*

1. *For any probability* $0 < u \leq \alpha \leq 1$*, the TCE or ES is given by the expression*

$$\mathrm{ES}_u[X] = \frac{\alpha}{u}\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma\left(\frac{2}{\lambda}\right)}{\Gamma\left(\frac{1}{\lambda}\right)}\left[1 - I\left(\frac{I^{-1}\left(1-\frac{u}{\alpha},\frac{1}{\lambda}\right)}{\sqrt{2}},\frac{2}{\lambda}\right)\right] + q_u, \tag{8.40}$$

where q_u *is the* u*-quantile of the loss distribution* X *with*

$$q_u = -\left[\frac{\alpha^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\frac{u}{\alpha},\frac{1}{\lambda}\right)\right]^{1/\lambda}. \tag{8.41}$$

2. *For any probability* $0 < \alpha < u \leq 1$*, the TCE or ES is given by the expression*

$$\begin{aligned}\mathrm{ES}_u[X] = {} & \frac{\alpha}{u}\frac{\alpha}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma\left(\frac{2}{\lambda}\right)}{\Gamma\left(\frac{1}{\lambda}\right)} \\ & - \frac{(1-\alpha)}{u}\frac{(1-\alpha)}{\delta_{\alpha,\lambda}^{1/\lambda}}\frac{\Gamma\left(\frac{2}{\lambda}\right)}{\Gamma\left(\frac{1}{\lambda}\right)} I\left(\frac{I^{-1}\left(1-\frac{1-u}{1-\alpha},\frac{1}{\lambda}\right)}{\sqrt{2}},\frac{2}{\lambda}\right) + q_u,\end{aligned} \tag{8.42}$$

where q_u *is the* u*-quantile of the loss distribution* X *with*

$$q_u = -\left[\frac{(1-\alpha)^\lambda}{\delta_{\alpha,\lambda}\sqrt{\lambda}}\right]^{1/\lambda}\left[I^{-1}\left(1-\frac{1-u}{1-\alpha},\frac{1}{\lambda}\right)\right]^{1/\lambda}. \tag{8.43}$$

Having discussed several families of loss models for which there are closed-form VaR and ES expressions available for the severity model, it is evident that there will be many loss models and severity distributions of interest that will not admit such closed-form expressions. Therefore, one may wish to consider alternative approaches such as specialized series expansion methods and empirical estimation of quantiles and tail expectations from the losses in each year. In the next section, we provide a brief overview of some results of this kind before discussing the SLA expansions that may be used in practice for compound process LDA models and then specialized series expansion results for heavy-tailed LDA loss models.

8.4 Non-Parametric Estimators for Quantile and Conditional Tail Expectation Functions

Marinelli *et al.* (2007) and Del Moral *et al.* (2013) consider empirical process estimators for the VaR and ES. When one does not wish to necessarily make a parametric model approximation, non-parameteric methods can be powerful tools that can be utilized within the SLA framework discussed below to develop expressions for the compound process VaR and ES. In particular, here we will discuss the application of such methods to estimation of the empirical quantile function and ES for the severity model, that is, losses in a given year, which allows one to avoid discarding information in an aggregation procedure if one were working on a non-parametric approximation at an annual loss level. To clarify this point, we note that one has two choices:

- Option 1 involves making a non-parametric estimate of the severity distribution and its tail functional (quantile function and ES) from the losses within each year and then form a point estimator from these few years of samples; in some cases, this would be the same as taking all the individual losses and estimating these empirical functionals from these losses (non-parametric estimate of the severity model in the LDA structure).

- Option 2 involves making a non-parametric estimate of the annual loss distribution and its tail functional (quantile function and ES) from the aggregated losses from each year.

As, typically in OpRisk, there are not a significant number of years of data, this second approach is deemed typically not viable for most loss processes; therefore, this section deals with the first option. Having made the empirical estimate of the severity model tail functionals, one can then combine this into the SLA structure to estimate the annual losses tail functionals for capital approximations. We discuss below some properties of estimation of the empirical quantile function and ES.

In addition, one may also be interested in a non-parametric estimation as such results may be useful as comparative analysis against an LDA parametric model approximation where a given parameter severity model is fitted to the loss data. A comparison of the parametric and non-parametric results is instructive about the associated model risk that one could encounter in estimating the capital under a parametric LDA model versus a purely non-parametric approach to calculation of the SLA for capital estimation.

In Cruz *et al.* (2015, chapter 8), we discussed empirical process estimations based on samples for goodness-of-fit testing; now we borrow from these discussions in the context of SLA developments for capital approximations. Therefore, instead of working solely with the empirical characteristic function and empirical distribution functions as we did in those chapters, we now work with the empirical quantile functions and empirical-sample-based estimators for the tail functionals such as ES. We first recall some results about the empirical estimation of the severity distribution function given for a sample size of n inter-annual i.i.d. losses by

$$\widehat{F}_X(x;n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[X_i \leq x], \tag{8.44}$$

with $X_i \sim F_X(x)$ where no assumptions on the parametric form or properties of $F_X(x)$ are made.

Note that this is precisely the same empirical process estimator discussed in Cruz *et al.* (2015, chapter 8) on model selection, when discussing omnibus tests for goodness-of-fit under the Kolmogorov-Smirnov (K-S) and Anderson-Darling (AD) tests. Recall that the above empirical process estimator has the property that it will satisfy almost sure convergence to the true distribution function according to the Glivenko–Cantelli theorem which means that one has the following almost sure convergence:

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\widehat{F}_X(x;n) - F_X(x)| = 0 \quad a.s. \tag{8.45}$$

Hence, as discussed in Del Moral *et al.* (2013), one may utilize such empirical process estimators for the loss distribution to estimate the quantile function according to the estimator based on the sample order statistics given by

$$\widehat{Q}_X(u,n) = \widehat{F}_X^{-1}(u,n) = X_{(i,n)}, \tag{8.46}$$

with $u \in \left(\frac{i-1}{n}, \frac{i}{n}\right]$ and order statistics for n samples given by

$$X_{(1,n)} = \min\{X_1, X_2, \ldots, X_n\} \leq X_{(2,n)} \leq \cdots \leq X_{(n,n)} = \max\{X_1, X_2, \ldots, X_n\}. \tag{8.47}$$

One can then show the following asymptotic approximation holds for the quantile function approximation above as given in Proposition 8.8, see discussion in Del Moral *et al.* (2013) and Marinelli *et al.* (2007).

Proposition 8.8 (Asymptotics of the Empirical Quantile Function Process) *Consider the estimated quantile function $\widehat{Q}_X(u,n)$ for any probability level $u \in (0,1)$ as $n \to \infty$. If the loss process satisfies that losses are distributed according to $X_i \overset{i.i.d.}{\sim} F_X(x)$ and this population distribution satisfies that $F_X(x)$ is continuously twice differentiable at the point $F_X^{-1}(u)$ and the density satisfies at this point that it is strictly positive, that is, $f_X(F_X^{-1}(u)) > 0$, then one has the following expansion:*

$$\sqrt{n}\left(\widehat{Q}_X(u,n) - Q_X(u)\right) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbb{I}[X_i \leq F_X^{-1}(u) - u]}{f_X(F_X^{-1}(u))} + o_u(1), \tag{8.48}$$

and the following central limit theorem (CLT) convergence in distribution applies as $n \to \infty$ for all $u \in (0,1)$

$$\sqrt{n}\left(\widehat{Q}_X(u,n) - Q_X(u)\right) \overset{d}{=} Normal\left(0, \frac{u(1-u)}{f_X^2\left(F_X^{-1}(u)\right)}\right). \tag{8.49}$$

In addition, if one assumes that the severity distribution $F_X(x)$ is in the class of continuous distributions with first derivative on the interval $[a,b]$ for some values of $a = F_X^{-1}(u_1) - \varepsilon$ and $b = F^{-1}(u_2) + \varepsilon$ for some $\varepsilon > 0$ and $f_X(x) > 0$ on interval $[a,b]$, that is, $F_X(x) \in C^1([a,b])$, then one has convergence of the empirical process for the quantile function to the Brownian bridge according to

$$\sqrt{n}\left(\widehat{Q}_X(u,n) - Q_X(u)\right) \to \frac{B_0}{f_X(F_X^{-1}(u))}, \tag{8.50}$$

with B_0 the Brownian bridge discussed in Cruz *et al.* (2015, chapter 8), see discussions in Marinelli *et al.* (2007).

One can also observe that the accuracy of such an empirical process estimator of the quantile function can be obtained via a confidence interval on the quantile point estimator via one of the two basic approaches:

1. Via the asymptotic normality result in Proposition 8.8 using the fact that as $n \to \infty$ for any desired $u \in (0,1)$

$$\sqrt{n}(\widehat{Q}_X(u,n) - Q_X(u)) \overset{d}{=} Normal\left(0, \frac{u(1-u)}{f_X^2\left(F_X^{-1}(u)\right)}\right). \tag{8.51}$$

   Hence, one can obtain the standard Gaussian distribution confidence interval so long as the asymptotic variance can be calculated exactly or approximately accurately, that is, one can evaluate the square of the severity density at the point $F_X^{-1}(u)$.

2. Alternatively, if the asymptotic variance is not accurately estimated or the number of samples is deemed too low for accurate asymptotic approximation under normality, then one may also obtain the confidence interval for the quantile point estimator using the U-statistics. Given data $\{X_i\}_{i=1}^n$, one can also define $\{U_i\}_{i=1}^n$ where $U_i = F(X_i)$ based on the hypothetical population model distribution $F(x)$ which produces independent uniform r.v.'s U_i. This produces the relationship

$$\Pr\left(X_{(k,n)} < F^{-1}(p) \leq X_{(l,n)}\right) = \Pr\left(U_{(k,n)} < p \leq U_{(l,n)}\right). \tag{8.52}$$

Then one may select indices k and l such that the following lower and upper confidence intervals hold:

$$\begin{aligned} \tfrac{k}{n} &= p - z_{\frac{r}{2}}\sqrt{\tfrac{p(1-p)}{n}}, \\ \tfrac{l}{n} &= p + z_{\frac{r}{2}}\sqrt{\tfrac{p(1-p)}{n}}, \end{aligned} \tag{8.53}$$

for a normal z-score at level $\frac{r}{2}$ where r corresponds to the level of significance of the following limit obtained by observing that the event $\{U_{(k,n)} < p \leq U_{(l,n)}\}$ is asymptotically equivalent in probability as $n \to \infty$ to the event of interest $\left\{\sqrt{n}\left|\widehat{F}_X^{-1}(u,n) - p\right| \leq z_{\frac{r}{2}}\sqrt{p(1-p)}\right\}$ such that they each occur asymptotically with probability

$$\begin{aligned} \lim_{n\to\infty} \Pr(U_{(k,n)} < p \leq U_{(l,n)}) &= \lim_{n\to\infty} \Pr\left(\sqrt{n}\left|\widehat{F}_X^{-1}(u,n) - p\right| \leq z_{\frac{r}{2}}\sqrt{p(1-p)}\right) \\ &= 1 - r. \end{aligned} \tag{8.54}$$

To empirically approximate the ES from the yearly loss data, one can trivially estimate this tail integral as follows for each year worth of losses and average the result, where for n losses in the year, one has the conditional tail expectation evaluated as

$$\widehat{\mathrm{ES}}[X] = \frac{1}{n-k}\sum_{i=k}^{n} X_{(i,n)}, \tag{8.55}$$

where $X_{(i,n)}$ is the i-th order statistic for the loss in a given year with n total losses and k corresponds to the quantile level at which ES is calculated with respect to, that is, the first index for the ordered losses such that $X_{(k,n)} > \widehat{\mathrm{VaR}}_\alpha[X]$. One can note that such an estimator is also consistent because of the law of large numbers.

Next, the above results developed for quantile and tail functionals of the severity model are combined within the LDA model structure for the compound loss process to obtain an SLA for capital measures. This will be considered under different assumptions on the tail properties of the severity model as the natural extension of the series expansion results on the tail of the compound process developed in Chapter 7.

8.5 First- and Second-Order SLA of the VaR for OpRisk LDA Models

This section considers the risk measure that is regulatory standard, the VaR. This risk measure has found prominence amongst practitioners as the risk measure of choice because of certain properties that it possess that make it efficient to evaluate. As noted in Dowd and Blake (2006), the VaR is a common measure of risk over different loss processes in the OpRisk context which can be universally applied to any risk type for any line of business. It allows for a fair comparison of risk across each risk type and business unit in an objective and quantitative manner.

In addition, VaR will allow one to aggregate the loss processes faced by a financial institution in such a manner that one may choose to take into account the ways in which each loss process may be correlated with each other. Another positive aspect of the VaR risk measure is that it is a probabilistic measure of risk which provides information on the probabilities associated with specified loss amounts.

It must also be said that there are some theoretical and practical downsides to considering VaR for the capital estimation, see discussion in Dowd and Blake (2006). One of the key practical limitations of the VaR risk measure when used for capital estimation is that the VaR does not consider tail events or values that exceed the VaR. This was described by Dowd and Blake (2006) as a negative because it only tells one the most that can be lost in 'good states' where an extreme loss from a tail event has not occurred. Unfortunately, it, therefore, does not provide any information on the magnitude of potential catastrophic losses in 'bad' states where a tail event does occur. From a theoretical perspective, it is also well known that VaR is not a coherent risk measure, see discussion in Cruz *et al.* (2015, chapter 6).

Having defined an asymptotic representation for the compound process for the OpRisk LDA single risk model, one can then obtain a convenient representation of the quantile function and, therefore, the VaR for any desired significance level. Such an analytic approximation is relevant to practitioners as it allows one to then evaluate risk measures such as the capital as defined by the VaR in closed form under this approximation. It was shown in Böcker and Klüppelberg (2005) that when considering such sub-exponential family loss process models, the VaR at a quantile level α, corresponding to a α quantile of the annual loss distribution $F_Z(z)$, could be approximated in closed form as given in Theorem 8.2 from Böcker and Klüppelberg (2005, theorem 2.4).

Theorem 8.2 (Single Loss Approximation VaR) *Consider an annual loss distribution $F_Z(z)$ under an LDA formulation, which satisfies Theorem 7.7 for which the annual loss distribution $F_Z(z)$ is represented according to an SLA given by*

$$\overline{F}_Z(x) = \mathbb{E}[N]\overline{F}_X(x)(1 + o(1)) \; as \; x \to \infty, \tag{8.56}$$

or, equivalently, the tail of the annual loss distribution is sub-exponential in behaviour with asymptotic equivalence,

$$\overline{F}_Z(x) \sim \mathbb{E}[N]\overline{F}_X(x).$$

Then the VaR for a level α is given by

$$\text{VaR}_\alpha[Z] := F_Z^{\leftarrow}(\alpha) = F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left[1 + o(1)\right]\right), \; as \; \alpha \to 1. \tag{8.57}$$

Here we define the notation used for the generalized inverse according to Definition 8.8.

Definition 8.8 (Generalized Inverse of a Distribution) *The generalized inverse of a distribution function $F(x)$, denoted by $F^{\leftarrow}(\alpha)$, is defined to be*

$$F^{\leftarrow}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \; \forall \alpha \in (0, 1). \tag{8.58}$$

To understand how this result is obtained, consider the result presented in Theorem 7.7 and note that clearly as $\alpha \to 1$, then this is equivalent to taking $x \to \infty$. Next, utilize the little-Oh Landau notation in which $o(1)$ represents a function tending to 0 as its argument tends to a boundary, that is, $\alpha \to 1$ or $x \to \infty$. Hence, this gives the representation

$$F_Z(x) = 1 - \mathbb{E}[N]\overline{F}_X(x)(1 + o(1)), \; x \to \infty. \tag{8.59}$$

Next consider setting the quantile level to α, that is, $F_Z(x) = \alpha$ and rearrange to obtain

$$F_X(x) = 1 - \frac{1-\alpha}{\mathbb{E}[N]}(1 + o(1)),\ x \to \infty, \tag{8.60}$$

which then produces the desired approximation for the SLA-VaR given by taking the inverse cumulative distribution function (CDF)

$$x = F_Z^{\leftarrow}(\alpha) = F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}\,[N]}(1 + o(1))\right),\ \alpha \to 1. \tag{8.61}$$

Next we discuss the fact that what we have detailed at this stage is the quantile approximation of the form

$$\mathrm{VaR}_\alpha[Z] := F_Z^{-1}(\alpha) = F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\,[1 + o(1)]\right),\ \text{as } \alpha \to 1 \tag{8.62}$$

defined with respect to the generally intractable distribution function for the annual loss random variable Z or with respect to the severity distribution generalized inverse evaluated with an argument of an asymptotic expansion which is of order $o(1)$. This result specified in Böcker and Klüppelberg (2005) is termed the closed-form OpRisk VaR approximation. However, it is clearly not the case that this asymptotic expansion will automatically be asymptotically equivalent to a function evaluation based on the severity distribution evaluated at the first term of the expansion with the remaining terms being of $o(1)$ such as the one given by

$$F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\,[1 + o(1)]\right) = F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right)[1 + o(1)],\ \text{as } \alpha \to 1, \tag{8.63}$$

unless some additional assumptions are satisfied. As noted in Degen (2010), in general, the result for the generalized inverse will not automatically imply that such a desired expansion in Equation 8.63, which is what we would like to achieve for the representation of the SLA, would be guaranteed.

To clarify, we would like to be able to represent the high quantiles, for high levels of α, from the intractable annual loss distribution by quantiles of the single-loss distribution severity model. In achieving this, we are able to circumvent the intractability of the annual loss distribution asymptotically and obtain in several cases an analytic VaR representation in terms of the parameters of the frequency distribution and severity distribution utilized in the LDA model.

To proceed with this approximation, in particular working with the result in Equation 8.63, we consider a sub-family of severity distributions in which this result will hold. In Degen (2010), it is detailed that if one would like to work with the resulting approximation for which this section takes its name 'single-loss' approximation, we must make some further restriction on the family of severity models considered. This will allow us to get the equivalent approximation for the high confidence level quantiles of the annual loss distribution in terms of those for the severity distribution, as detailed in Equation 8.63. This is achieved by considering a restricted sub-class of distributions within in the sub-exponential family that was discussed in Chapter 3. This restricted sub-family of distribution will be denoted by $\mathcal{S} \in \mathscr{F}$. Selection of this sub-class turns out to be sufficiently broad so as to be of interest to several models that practitioners modeling OpRisk LDA models may utilize in practice and have the additional feature that they satisfy the SLA presented in Definition 8.9.

Definition 8.9 (Single-Loss Approximation VaR in terms of Severity Quantile Function) *Assuming that the result relating the annual loss distribution F_Z to the severity distribution F_X given by*

$$F_Z^{\leftarrow}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\left[1+o(1)\right]\right)=F_X^{-1}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\right)\left[1+o(1)\right],\ as\ \alpha\to 1 \tag{8.64}$$

holds for the class of severity distributions under consideration, then the SLA will be given by

$$F_Z^{-1}(\alpha)\approx F_X^{-1}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\right). \tag{8.65}$$

Remark 8.6 *The SLA-VaR approximation provided in Definition 8.9 should be used with caution; it has been noted in Mignola and Ugoccioni (2006), Böcker and Sprittulla (2006), Degen (2010) and Peters* et al. *(2013) that such an approximation may result in a large approximation error and this error will be a function of the quantile level α for which the VaR is calculated as well as the family of severity models considered and its parameter values selected for the severity and frequency model in the LDA structure.*

It is interesting to note the interpretation of this result provided in Böcker and Sprittulla (2006). This is developed by the authors according to the following arguments where they first assume that the distributions for the severities is a member of the sub-exponential family; therefore, one has that

$$\mathrm{VaR}_\alpha[Z]:=F_Z^{-1}(\alpha)=F_Z^{\leftarrow}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\left[1+o(1)\right]\right),\ \text{as}\ \alpha\to 1. \tag{8.66}$$

This result is based on the fact that for sub-exponential family members, the relationship between the tail distribution of the annual loss distribution $\overline{F}_Z(z)$ and the tail of the severity distribution $\overline{F}_X(z)$ satisfy the relation $\overline{F}_Z(z)\sim\mathbb{E}[N]\overline{F}_X(z)$ as $z\to\infty$. Next recall the representation of the tail of the annual loss distribution in Theorem 7.7 (Equation 7.99) and consider a result from Embrechts *et al.* (1997, definition 1.3.3) which shows that when considering sub-exponential family members of severity distribution, one can show that $\overline{F}_X^{(n)*}(z)\sim n\overline{F}_X(z)$ as $z\to\infty$. Using this representation and this result of Embrechts *et al.* (1997), Böcker and Sprittulla (2006) made the substitution to obtain

$$\mathbb{P}\mathrm{r}\left(\sum_{s=1}^{n}X_s>z|N=n\right)=\overline{F}_X^{(n)*}(z)\sim n\overline{F}_X(z)\sim n\overline{F}_X(z)F_X^{(n-1)*}(z). \tag{8.67}$$

Then defining $p=\overline{F}_X(z)$ and $(1-p)=F_X^{(n-1)*}(z)$, they presented an interpretation of the SLA-VaR approximation according to a binomial probability model. In particular, consider a Bernoulli random variable for which the probability of success is denoted by the defined probability p. The probability of success in this scenario, therefore, represents the situation in which a large loss that impacts the VaR significantly is experienced and occurs rarely with probability $p=\overline{F}_X(z)$. Conversely, the probability of failure, defined by $(1-p)=F_XF_X^{(n-1)*}(z)$, represents a small or negligible loss with respect to altering the VaR of the annual loss. Under this interpretation, one may consider the asymptotic ($z\to\infty$) interpretation of the conditional probability $\mathbb{P}\mathrm{r}\left(\sum_{s=1}^{n}X_s>z|N=n\right)$ as represented

approximately by a binomial distribution with the number of successes in n trials equal to 1, which recovers the SLA for OpRisk VaR.

Remark 8.7 *A consequence that is apparent from this interpretation of the SLA-VaR approximation provided in Definition 8.9 is that this approximation will typically underestimate OpRisk VaR because, typically, there may be more than a single loss that will contribute substantially to the annual loss process in a given year.*

Remark 8.8 *A second consequence that is apparent from this interpretation of the SLA-VaR approximation provided in Definition 8.9 is that it would be analogous to truncating the exact closed-form expressions obtained in Chapter 7 after one loss.*

In the following section, we provide some examples for which such SLAs of compound process LDA models admit analytic closed-form expressions for popular two-parameter severity distribution models comprising sub-families of the sub-exponential family satisfying their membership of $\mathcal{S}$. As noted in Böcker and Klüppelberg (2005) the fact that regulatory capital and economic capital measurements for OpRisk are based on very high quantiles of the annual loss distribution $F_Z(z)$, then the quantification of such risk measures can be sufficiently estimated via the SLAs for a family of Pareto severity models as detailed in Examples 8.3 and 8.2 summarized below, see derivations in Böcker and Klüppelberg (2005).

EXAMPLE 8.2 Poisson-Weibull LDA Model SLA-VaR

Consider the Weibull LDA model constructed via a severity distribution model in which losses X_i in the tth year are assumed to be distributed according to a Weibull distribution with shape and scale parameters denoted by $\tau > 0$ and $\theta > 0$, respectively. Furthermore, we assume that the frequency distribution considered will be Poisson with intensity parameter λ for the mean number of counts $\mathbb{E}[N_t]$ in the t-th year. Under this choice of severity model, one can write down analytically according to Definition 2.3 the tail of the severity distribution given by

$$\overline{F}_X(x) = \exp\left(-\left(\frac{x}{\theta}\right)^{\tau}\right), \ x > 0.$$

Substituting this result into the SLA, one can obtain an analytic expression for the annual loss $\mathrm{VaR}_\alpha[Z]$ as follows:

$$\mathrm{VaR}_\alpha[Z] = \theta\left[\ln\left(\frac{\lambda}{1-\alpha}\right)\right]^{1/\tau}, \ \alpha \to 1.$$

A plot of the tail approximation obtained from a Monte Carlo simulation with a million-generated annual years and empirical estimation of the tail quantiles versus the SLA evaluated over the same range of levels $\alpha \in [0.8, 0.9999]$ for a range of different severity parameter combinations (Figure 8.1).

The results from this analysis show that as λ increases, that is, the average number of losses in a year increases, the accuracy of the first-order SLA for the quantiles degrades. This demonstrates that it is important to consider the accuracy of such approximations for given parameter ranges as well as quantile levels.

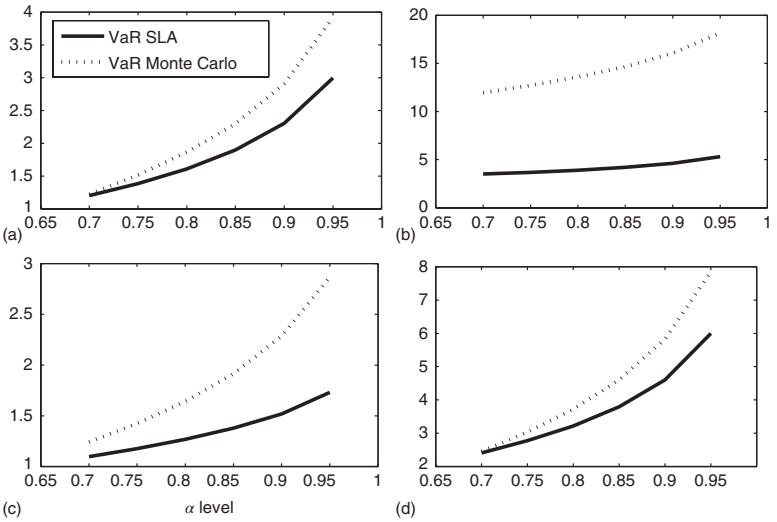

FIGURE 8.1 **Comparison of Monte Carlo annual loss VaR and SLA for the Poisson-Weibull LDA model. (a) Case 1: Poisson $\lambda = 1$ and Weibull shape and scale $(\theta = 1, \tau = 1)$; (b) Case 2: Poisson $\lambda = 10$ and Weibull shape and scale $(\theta = 1, \tau = 1)$; (c) Case 3: Poisson $\lambda = 1$ and Weibull shape and scale $(\theta = 2, \tau = 1)$; and (d) Case 4: Poisson $\lambda = 1$ and Weibull shape and scale $(\theta = 1, \tau = 2)$.**

EXAMPLE 8.3 Poisson-Pareto-LDA Model SLA-VaR

Consider the Pareto-LDA model constructed via a severity distribution model in which losses X_i in the tth year are assumed to be distributed according to a two-parameter Pareto distribution in which the shape and scale parameters are restricted to be strictly positive $a > 0$ and $\theta > 0$. Furthermore, we assume that the frequency distribution considered will be Poisson with intensity parameter λ for the mean number of counts $\mathbb{E}[N_t]$ in the t-th year. Under this choice of severity model, one can write down analytically according to Definition 2.3 the tail of the severity distribution given by

$$\overline{F}_X(x) = \left(1 + \frac{x}{\theta}\right)^{-a}, \; x > 0.$$

Substituting this result into the SLA, one can obtain an analytic expression for the annual loss $\mathrm{VaR}_\alpha[Z]$ as follows:

$$\mathrm{VaR}_\alpha[Z] \sim \theta\left(\frac{\lambda}{1-\alpha}\right)^{1/a}.$$

Hence, we can consider the analytic approximation given by

$$\mathrm{VaR}_\alpha[Z] = \theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1\right], \; \alpha \to 1.$$

To conclude the discussion on the standard SLA-VaR approximation, we make the following comment relating to time scaling properties of the SLA-VaR approximation in the case of the Poisson-Pareto-LDA model SLA-VaR approximation, pointed out in Böcker and Klüppelberg (2005).

Corollary 8.1 (Time Scaling Property of SLA-VaR) *For the single risk LDA models with slowly varying severity models in which* $\overline{F}_X(x) \sim (x/\theta)^{-a}$ *as* $x \to \infty$*, such as the Pareto severity model, then the resulting SLA-VaR expression suggests that if one considers one of the two cases:*

- ***Case 1.*** *Poisson process for the number of losses over time given by* $(N_t)_{t\geq 0}$ *with* $\lambda > 0$ *such that*

$$\mathbb{P}\mathrm{r}(N(t) = n) = \exp(-\lambda t)\frac{(\lambda t)^n}{n!},\ n \in \mathbb{J}^+, \tag{8.68}$$

  then the result for the SLA-VaR approximation is given by

$$\mathrm{VaR}_\alpha[Z(t)] = \theta\left[\left(\frac{\lambda t}{1-\alpha}\right)^{1/a} - 1\right],\ \alpha \to 1.$$

- ***Case 2.*** *Negative binomial process accounting for overdispersion in the count process for the number of losses over time given by* $(N_t)_{t\geq 0}$ *with* $\beta, \gamma > 0$ *such that*

$$\mathbb{P}\mathrm{r}(N(t) = n) = C_n^{\gamma+n-1}\left(\frac{\beta}{\beta+1}\right)^{\gamma}\left(\frac{t}{\beta+t}\right)^{n},\ n \in \mathbb{J}^+, \tag{8.69}$$

  then the result for the SLA-VaR approximation is given by

$$\mathrm{VaR}_\alpha[Z(t)] = \theta\left[\left(\frac{\gamma t}{\beta(1-\alpha)}\right)^{1/a} - 1\right],\ \alpha \to 1.$$

Furthermore, in both the cases, the VaR in period t *can be represented as a time-scaled version in period 1, according to the following* a*-root time law for operational value at risk (OpVaR),*

$$\mathrm{VaR}_\alpha[Z(t)] \sim t^{1/a}\mathrm{VaR}_\alpha[Z],\ \alpha \to 1. \tag{8.70}$$

8.5.1 SECOND-ORDER REFINEMENTS OF THE SLA VaR FOR HEAVY-TAILED LDA MODELS

Refinements of the SLA-VaR approximation in Definition 8.9 have been developed by several authors, for example, Böcker and Sprittulla (2006), Degen (2010) and Peters *et al.* (2011). In the following, we summarize a few of these refinements of particular relevance to practitioners in OpRisk, providing some examples of SLA-VaR approximations under these refinements for the working examples detailed previously in Example 8.3.

On the basis of the results of Proposition 7.15, one may obtain the following general refined 'second-order' asymptotic expansion for the SLA given in Theorem 8.3. These results are a direct consequence of the properties of second-order sub-exponentiality, see discussion in Teugels (1975). Note that these second-order results involve expanding the argument of the series expansion for a second term and are directly based on the extensive discussions of tail expansions developed in Chapter 7.

Theorem 8.3 (Refined Second-Order VaR Single-Loss Approximations) *Assume that the severity distribution satisfies the conditions in Proposition 7.15 and if the mean loss is finite, then assume further that the hazard rate* $h(x) = \frac{f_X(x)}{1-F_X(x)}$ *is of regular variation with* $h \in RV_{-\beta}$ *for* $\beta \geq 0$*, then as* $\alpha \to 1$*, one has*

$$F_Z^{-1}(\alpha) = F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left\{1 + \widetilde{c}_\beta g_1\left(F_X^{-1}(\widetilde{\alpha})\right) + o(g_1(F_X^{-1}(\widetilde{\alpha})))\right\}^{-1}\right), \tag{8.71}$$

where $\widetilde{\alpha} = 1 - (1-\alpha)/\mathbb{E}[N]$ *and*

$$\begin{aligned} g_1(x) &= \begin{cases} \dfrac{f_X(x)}{1-F_X(x)}, & \text{if } \mathbb{E}[X] < \infty, \\ \dfrac{\int_0^x \overline{F}_X(s)ds f_X(x)}{1-F_X(x)}, & \text{if } \mathbb{E}[X] = \infty. \end{cases} \\ \widetilde{c}_\beta &= \begin{cases} \dfrac{\mathbb{E}[X]\mathbb{E}[(N-1)N]}{\mathbb{E}[N]}, & \text{if } \mathbb{E}[N] < \infty, \\ \dfrac{c_\beta \mathbb{E}[(N-1)N]}{\mathbb{E}[N]}, & \text{if } \mathbb{E}[N] = \infty. \end{cases} \end{aligned} \tag{8.72}$$

In understanding this second-order SLA-VaR approximation, we start by considering the refinement provided in Böcker and Sprittulla (2006) in which they develop the modified SLA-VaR approximation given in Proposition 8.9. These authors observed that the accuracy of the first-order SLA-VaR degraded especially in scenarios in which the expected loss in a given year is non-negligible relative to the VaR. One situation in which this may occur is the scenario in which the mean frequency of losses in a given year $\lambda = \mathbb{E}[N]$ is large and the typical losses experienced are not extremely heavy tailed; in this scenario, the SLA-VaR in Definition 8.9 can have poor accuracy unless α is very close to 1. However, in this particular chapter, this will not be a particularly interesting scenario; we do, however, provide the modification to the SLA-VaR provided to accommodate an improvement in the accuracy of the SLA-VaR for such a scenario.

As discussed in Böcker and Sprittulla (2006), the justification provided for this adjustment is based on an argument of centring because of the theory of large deviations. Given the property of sub-exponential distributions discussed in Embrechts *et al.* (1997, lemma 1.3.5) which shows that

$$\lim_{x\to\infty} \frac{\overline{F}_X(x-y)}{\overline{F}_X(x)} = 1, \ y \in \mathbb{R}, \tag{8.73}$$

then when this is considered in the context of the LDA models discussed, it implies that

$$\begin{aligned} \overline{F}_Z(x) &\sim \overline{F}_Z(x+\lambda\mu), \ x \to \infty \\ \overline{F}_X(x) &\sim \overline{F}_X(x+\mu), \ x \to \infty. \end{aligned} \tag{8.74}$$

When this is combined with Theorem 7.7, one obtains the result derived in Böcker and Sprittulla (2006) and summarized in Proposition 8.9.

Proposition 8.9 (Single-Loss Approximation Refinement by Mean Correction Böcker and Sprittulla (2006)) *Assume that the severity distribution is both the member of the sub-exponential*

family of distributions $F_X(x) \in \mathcal{F}$ and has finite mean $\mu = \mathbb{E}[X_i] < \infty$. Consider the transformed annual loss random variable involving a centring according to the mean annual loss, defined by $\widehat{Z}_t = Z_t - \mathbb{E}[N_t]\mathbb{E}[X_i(t)]$. According to Theorem 8.2, one obtains the result for the annual loss distribution given by

$$\overline{F}_Z(x + \mathbb{E}[N]\mathbb{E}[X]) \sim \mathbb{E}[N]\overline{F}_X(x + \mathbb{E}[X]), \text{ as } z \to \infty. \tag{8.75}$$

This result then provides the refinement of the SLA-VaR in Definition 8.9 given by

$$\widetilde{\mathrm{VaR}}_\alpha[Z] = F_X^{\leftarrow}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left[1 + o(1)\right]\right) + (\mathbb{E}[N] - 1)\mathbb{E}[X], \text{ as } \alpha \to 1. \tag{8.76}$$

Therefore, the refined SLA-VaR that can be approximately evaluated by

$$\widetilde{\mathrm{VaR}}_\alpha[Z] \approx F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right) + (\mathbb{E}[N] - 1)\mathbb{E}[X], \text{ as } \alpha \to 1. \tag{8.77}$$

Remark 8.9 *The two components of this refinement to the OpRisk LDA SLA-VaR approximation can be interpreted directly in terms of the loss process. The first term corresponds to a single very high consequence extreme loss at a high confidence level and the second term corresponds to $(n-1)$ expected losses of expected loss size.*

Remark 8.10 *As discussed in Böcker and Sprittulla (2006), because this constant mean correction term is independent of the level at which the tail approximation is obtained as $\alpha \to 1$, this correction of $(\mathbb{E}[N] - 1)\mathbb{E}[X]$ is not unique and any constant multiplier of this correction would also preserve the limiting behaviour. The actual choice selected was instead motivated by uniform convergence properties known to exist for sequences of centred random variates in large deviations theory.*

In Example 8.4, a comparison between the standard SLA-VaR approximation is compared to the mean-corrected refinement for cases in which the tail exponent of the Pareto model involves an EVI (tail index parameter) that produces a finite mean for the severity model. In addition, we consider two scenarios: the first involves cases in which the compound process average loss is insignificant relative to the VaR and in the second case the opposite situation in which the average annual loss is significant relative to the VaR. These two cases will aid in illustrating the significant impact such an approximation refinement can have in certain scenarios on the estimate of the VaR. We illustrate this for a spectrum of significance levels to represent regulatory standards under Basel II/III and internal economic capital reporting standards for financial institutions.

EXAMPLE 8.4 Poisson-Pareto-LDA Model SLA-VaR with Mean Correction

Consider the Pareto-LDA model constructed via a severity distribution model in which losses X_i in the tth year are assumed to be distributed according to a two-parameter Pareto distribution in which the shape and scale parameters are restricted to be strictly positive such that the mean of the Pareto severity distribution is finite, $a > 1$ and $\theta > 0$. Furthermore, we assume that the frequency distribution considered will be Poisson with intensity parameter λ for the mean number of counts $\mathbb{E}[N_t]$ in the t-th year. Under this choice of severity model, one can write down analytically according to Definition 2.3 the tail of the severity distribution given by

$$\overline{F}_X(x) = \left(1 + \frac{x}{\theta}\right)^{-a}, \ x > 0.$$

Substituting this result into the SLA, one can obtain an analytic expression for the annual loss $\mathrm{VaR}_\alpha[Z]$ as follows:

$$\mathrm{VaR}_\alpha[Z] \sim \theta\left(\frac{\lambda}{1-\alpha}\right)^{1/a}.$$

Hence, we can consider the analytic approximation given by the standard SLA-VaR

$$\mathrm{VaR}_\alpha[Z] = \theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{\frac{1}{a}} - 1\right],\ \alpha \to 1,$$

which will be compared to the refinement corresponding to the mean-corrected SLA-VaR

$$\widetilde{\mathrm{VaR}}_\alpha[Z] = \theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1\right] + (\lambda - 1)\frac{1}{\theta(a-1)},\ \alpha \to 1. \tag{8.78}$$

■

The impact of the mean correction when applied to the refinement of the SLA-VaR can be significant as illustrated in the example above. However, there are further refinements one can make to such approximations as pointed out in Degen (2010).

Remark 8.11 *It was observed in Degen (2010) that this refined mean-corrected approximation involves a simplification of the expression by dropping the expression corresponding to terms of order* $o(1)$ *as follows:*

$$F_Z^{-1}(\alpha) = F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left[1 + o(1)\right]\right)\left(1 + \frac{(\mathbb{E}[N]-1)\mathbb{E}[X]}{F_X^{-1}\left(1 - \frac{(1-\alpha)}{\mathbb{E}[N]}(1+o(1))\right)}\right). \tag{8.79}$$

where $F_X^{-1}\left(1 - \frac{(1-\alpha)}{\mathbb{E}[N]}(1+o(1))\right) = o(1)$. *Hence, Degen (2010) argued that the refined SLA-VaR approximation obtained in Equation 8.76 does not correspond in general with the approximation given in Equation 8.77.*

As a consequence of this observation, further refinements to the SLA-VaR have been proposed in the literature which extend these results. These are presented in the following two cases:

- ***Case I***: involving LDA models with heavy-tailed severity distributions with finite mean severity;
- ***Case II***: involving LDA models with heavy-tailed severity distributions with infinite mean severity.

The following sets of refined SLA-VaR approximation results are derived by considering the relative error in the SLA-VaR between the tail asymptotics based on the distribution for the severity model and the tail asymptotics based on the annual loss distribution. In particular, the relationships provided in Proposition 7.15 and subsequently the representation in Theorem 8.3 are of direct utility in deriving the refinements, see Omey and Willekens (1986) and Albrecher *et al.* (2010).

Proposition 8.10 (Single-Loss Approximation Refinement Degen (2010) Case I: Finite Mean) *Assume the density of the severity distribution is regularly varying with parameter β. That is, consider $\beta > 0$ and assume that the density of the severity distribution $f_X(x) \in RV_{-1/\beta-1}$ and, therefore, that the tail distribution of the severity in the LDA model satisfies $\overline{F}_X \in RV_{-1/\beta}$. In this scenario, it can be shown by using the expression for a second-order sub-exponential form and considering $g_1(x) = \frac{f_X(x)}{1-F_X(x)} \sim 1/(\beta x)$, as $x \to \infty$, that a refined SLA-VaR approximation is obtained according to the following expression:*

$$\widetilde{\mathrm{VaR}}_\alpha[Z] \approx F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right) + \left(\mathbb{E}[N] + \frac{\mathbb{V}\mathrm{ar}[N]}{\mathbb{E}[N]} - 1\right)\mathbb{E}[X], \ \textit{as } \alpha \to 1. \quad (8.80)$$

Note that in the case in which the index of regular variation is $\zeta = 0$, the concept of local heavy tailedness is proposed to obtain an approximation for the SLA-VaR.

Remark 8.12 *In the case that the frequency distribution of the LDA model is Poisson, the mean and variance of the number of observed losses in a given year are by definition equal, resulting in a dispersion of $\frac{\mathbb{V}\mathrm{ar}(N)}{\mathbb{E}(N)} = 1$. In this case, we see that the SLA-VaR approximation refinement of Case I in Degen (2010) involves a factor of $\mathbb{E}[N]$ whilst the approximation presented in Böcker and Spritzulla (2006) presented in Definition 8.9 instead involves a factor of $(\mathbb{E}[N] - 1)$. Furthermore, in the case of a negative binomial severity distribution in which there is overdispersion, this SLA-VaR approximation will differ by the amount of overdispersion present.*

In Example 8.5, we consider the illustration developed in Degen (2010) of the refined SLA-VaR result and compare the performance in several simulations to those of the standard SLA-VaR approximation developed in Definition 8.9.

EXAMPLE 8.5 SLA-VaR for Poisson-Pareto-LDA: Standard versus Refinement Case I

Consider the Pareto-LDA model of Example 8.3 in which losses X_i in the tth year are assumed to be distributed according to a two-parameter Pareto distribution in which the shape and scale parameters are restricted to be strictly positive $a > 0$ and $\theta > 0$. Furthermore, we assume that the frequency distribution considered will be Poisson with intensity parameter λ for the mean number of counts $\mathbb{E}[N_t]$ in the tth year. Under this model, one can derive the refined SLA-VaR from Definition 8.10 in contrast to that presented earlier. In particular, we consider the severity distribution given by a separation into a component for the body of the severity distribution $F_X(x) = F_{\mathrm{body}}(x)$ when the loss amount is between $0 \leq x \leq u$ and given for the extreme losses in which for $u \leq x \leq \infty$, we have the severity distribution given by

$$F_X(x) = 1 - \overline{F}_{\mathrm{body}}(u)\left(1 + \frac{(x-u)\zeta}{\theta}\right)^{-1/\zeta}. \quad (8.81)$$

Next to evaluate the SLA-VaR approximation, we need to obtain an expression for the inverse distribution function of the severity model. Therefore, at the relevant quantile level to calculate the VaR approximation such that $\alpha \geq F_X(u)$, one obtains the inverse distribution at level α for this severity model given by

$$F_X^{-1}(\alpha) = \frac{((1-\alpha)/\overline{F}_{\mathrm{body}}(u))^{-\zeta} - 1}{\zeta}\theta + u. \quad (8.82)$$

Hence, according to refined SLA-VaR result for the heavy-tailed and finite mean severity model in Definition 8.10, one obtains the approximation after substitution for this model, given by

$$\begin{aligned}\widetilde{\mathrm{VaR}}_\alpha[Z] &\approx F_X^{-1}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\right)+\left(\mathbb{E}[N]+\frac{\mathbb{V}\mathrm{ar}(N)}{\mathbb{E}(N)}-1\right)\mathbb{E}[X]\\ &=\frac{\theta}{\zeta}\left(\frac{(1-\alpha)}{\mathbb{E}[N]\overline{F}_{\mathrm{body}}(u)}\right)^{-\zeta}-\frac{\theta}{\zeta}+u+c,\ \text{as } F(u)\leq\alpha,\end{aligned}\tag{8.83}$$

where, in this case, $\mathbb{E}[X]<\infty$ and $c=\frac{\mathbb{E}[X]\mathbb{E}[(N-1)N]}{\mathbb{E}[N]}$. Alternatively, under the SLA-VaR approximation of Sahay *et al.* (2007), one obtains the approximation in this example given by

$$\begin{aligned}\widetilde{\mathrm{VaR}}_\alpha[Z] \approx{}& \frac{\theta}{\zeta}\left(\frac{1-\alpha}{\mathbb{E}[N]\overline{F}_{\mathrm{body}}(u)}\right)^{-\zeta}\\ &\times\left(1+\zeta\frac{\mathbb{E}[X]\mathbb{E}[(N-1)N]}{\theta\mathbb{E}[N]}\left(\frac{1-\alpha}{\mathbb{E}[N]\overline{F}_{\mathrm{body}}(u)}\right)^{\zeta}\right.\\ &\left.-\frac{(\zeta+1)\zeta}{2}\left(\frac{\mathbb{E}[X]\mathbb{E}[(N-1)N]}{\theta\mathbb{E}[N]}\right)^2\left(\frac{1-\alpha}{\mathbb{E}[N]\overline{F}_{\mathrm{body}}(u)}\right)^{2\zeta}\right)-\frac{\theta}{\zeta}+u.\end{aligned}\tag{8.84}$$

As we consider in this chapter the situations in which the severity distribution can have extremely heavy tails, this means that it is also sensible to consider the scenarios in which the mean annual loss can be infinite. The refined SLA-VaR approximation for this scenario has also been studied and the results are presented below.

Proposition 8.11 (Single-Loss Approximation Refinement Degen (2010) Case II: Infinite Mean) *Consider a severity distribution in the LDA model in which the average loss in a given year is infinite* $\mathbb{E}[X_i(t)]=\infty$. *Consider the case in which* $\beta>1$, *then it can be shown that assuming the expansion in Theorem 8.3 is satisfied such that*

$$g_1(x)\sim\frac{1}{\beta-1}\overline{F}_X(x),\ \textit{as } x\to\infty.\tag{8.85}$$

In this scenario, it can be shown by using second-order sub-exponential representation that a refined SLA-VaR approximation is obtained according to the following expression:

$$\begin{aligned}\widetilde{\mathrm{VaR}}_\alpha[Z] \approx{}& F_X^{-1}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\right)-(1-\alpha)F_X^{-1}\left(1-\frac{1-\alpha}{\mathbb{E}[N]}\right)\frac{c_\beta}{1-1/\beta}\\ &\times\left(1+\frac{\frac{\mathbb{V}\mathrm{ar}[N]}{\mathbb{E}[N]}-1}{\mathbb{E}[N]}\right),\ \textit{as } \alpha\to1\end{aligned}\tag{8.86}$$

with

$$c_\beta = (1-\beta)\frac{\Gamma^2(1-1/\beta)}{2\Gamma(1-2/\beta)}, \tag{8.87}$$

if $1 < \beta < \infty$. *In the case in which* $\beta = 1$, *one can show that* $c_\beta = 1$ *and the resulting refined SLA-VaR approximation reduces to*

$$\widetilde{\mathrm{VaR}}_\alpha[Z] \approx F_X^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right) + c_\beta \mathbb{E}[F^{-1}(\widetilde{\alpha})], \ \textit{as}\ \alpha \to 1 \tag{8.88}$$

with $\widetilde{\alpha} = 1 - (1-\alpha)/\mathbb{E}[N]$.

Remark 8.13 *It is clear that in Case II which involves an LDA model with severity distribution with infinite mean, the adjustment factor for the refined SLA-VaR is now a function of the quantile level* α.

An example comparison between the refined SLA-VaR in the Case II involving severity models in the LDA framework with infinite average loss will be studied with respect to an alternative approximations in the case of α-stable models in Section 4, based on the work in 2011. In particular, below, we present two detailed examples for the LDA models comprised Poisson-LogNormal and Poisson-IG, where we compare the first- and second-order VaR approximations.

EXAMPLE 8.6 Poisson-LogNormal LDA Model SLA-VaR

Consider the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N_t}$, and chosen to be a LogNormal distribution $X_i \sim F_X(x) = LogNormal(\mu, \sigma)$ where the two parameters in this model correspond to parameterizing the shape of the distribution for the severity σ and the log-scale of the distribution μ. The survival and quantile functions of the severity are given by

$$\begin{aligned} f(x;\mu,\sigma) &= \frac{1}{x\sqrt{2\pi\sigma^2}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},\ x>0;\ \mu\in\mathbb{R},\ \sigma>0, \\ \overline{F}(x;\mu,\sigma) &= 1 - F(x) = \int_x^\infty \frac{1}{u\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(\ln u - \mu)^2\right) du \\ &= \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{\ln x - \mu}{\sqrt{2\sigma^2}}\right],\ x>0;\ \mu\in\mathbb{R},\ \sigma>0, \end{aligned}$$

$$Q(p) = \exp(\mu + \sigma\Phi^{-1}(p)),\ 0<p<1.$$

Therefore, the closed-form SLA for the VaR risk measure at level α would be presented in this case under first- and second-order approximations for the annual loss $Z = \sum_{n=1}^N X_i$ according to Equations 8.89 and 8.90, respectively,

$$\mathrm{VaR}_\alpha[Z] = \exp\left[\mu - \sigma\Phi^{-1}\left(\frac{1-\alpha}{\lambda}\right)\right], \tag{8.89}$$

$$\mathrm{VaR}_\alpha[Z] = F_X^{-1}\left(1 - \frac{1-\alpha}{\lambda}\{1 + \widetilde{c}_\beta g_1(F_X^{-1}(\widetilde{\alpha}))\}^{-1}\right), \tag{8.90}$$

where $\widetilde{\alpha} = 1 - (1-\alpha)/\lambda$, $g_1(x) = \frac{f_X(x)}{1-F_X(x)}$ and $\widetilde{c}_\beta = e^{\mu+\sigma^2/2}\lambda$.

We will now compare these first-order and second-order asymptotic results to the crude Monte Carlo approach (for which one can generate uncertainty measures such as confidence intervals in the point estimator), the results are shown in Figure 8.2. To complete this example, we illustrate the basic Monte Carlo solution for the VaR for a range of quantile levels of the annual loss distribution; we display these along with the measured confidence intervals in the point estimators and compare these to the first-order SLA asymptotic result. The quantiles $\alpha \in \{0.70, 0.75, 0.80, 0.85, 0.9, 0.95, 0.99, 0.995, 0.9995\}$ are considered where the 99.5% and 99.95% quantile levels do in fact correspond to regulatory standards of reporting in Basel II/III. ■

Remark 8.14 *In the Poisson-LogNormal example, we can see in Figure 8.2 that the volatility of the severities, σ, plays a very important role on the accuracy of the approximations. It is also important to note that although in all the cases the second- order approximations does not perform very well for quantiles below the 95-th percentile, above this threshold, it usually performs better than the first order.*

EXAMPLE 8.7 Single Risk LDA Poisson-Inverse-Gaussian Family

Consider the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N_t}$, and chosen to be an Inverse Gaussian distribution $X_i \sim InverseGaussian(\widetilde{\mu}, \widetilde{\lambda})$ where the two parameters in this model correspond to parameterizing the mean of the distribution for the severity $\widetilde{\mu}$ and the shape of the distribution $\widetilde{\lambda}$. The survival and quantile functions of the severity are given by

$$f(x;\widetilde{\mu},\widetilde{\lambda}) = \left(\frac{\widetilde{\lambda}}{2\pi}\right)^{1/2} x^{-3/2} \exp\left\{-\frac{\widetilde{\lambda}(x-\widetilde{\mu})^2}{2\widetilde{\mu}^2 x}\right\},$$

$$F(x;\widetilde{\mu},\widetilde{\lambda}) = \Phi\left(\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}} - 1\right)\right) + \exp\left(\frac{2\widetilde{\lambda}}{\widetilde{\mu}}\right)\Phi\left(-\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}} + 1\right)\right).$$

Therefore, the closed-form SLA for the VaR risk measure at level α would be presented in this case under first- and second-order approximations for the annual loss $Z = \sum_{n=1}^{N} X_i$ according to Equations 8.91 and 8.92, respectively,

$$\mathrm{VaR}_\alpha[Z] = F_X^{-1}\left(\frac{1-\alpha}{\lambda}\right), \tag{8.91}$$

$$\mathrm{VaR}_\alpha[Z] = F_X^{-1}\left(1 - \frac{1-\alpha}{\lambda}\{1 + \widetilde{c}_\beta g_1(F_X^{-1}(\widetilde{\alpha}))\}^{-1}\right), \tag{8.92}$$

where $\widetilde{\alpha} = 1 - (1-\alpha)/\lambda$, $g_1(x) = \frac{f_X(x)}{1-F-X(x)}$ and $\widetilde{c}_\beta = \widetilde{\mu}\lambda$.

As we have seen in the closed-form LDA models in Chapter 6 that the IG family is closed under convolution, that is,

$$X_i \sim InverseGaussian(\widetilde{\mu}, \widetilde{\lambda}) \Rightarrow S_n = \sum_{i=1}^{n} X_i \sim InverseGaussian(n\widetilde{\mu}, n^2\widetilde{\lambda}),$$

we can calculate the distribution of the compound process analytically (see the comparison with the approximations in Figure 8.3). The drawback of this family is that there is no closed form for the inverse CDF, which obliges us to resort to a numerical procedure for obtaining the quantiles; fortunately, this is efficient and accurate for this class of models. For this model, we also present first- and second-order approximations for the VaR and SRM for different choices of parameters in Figures 8.4 and 8.5. ■

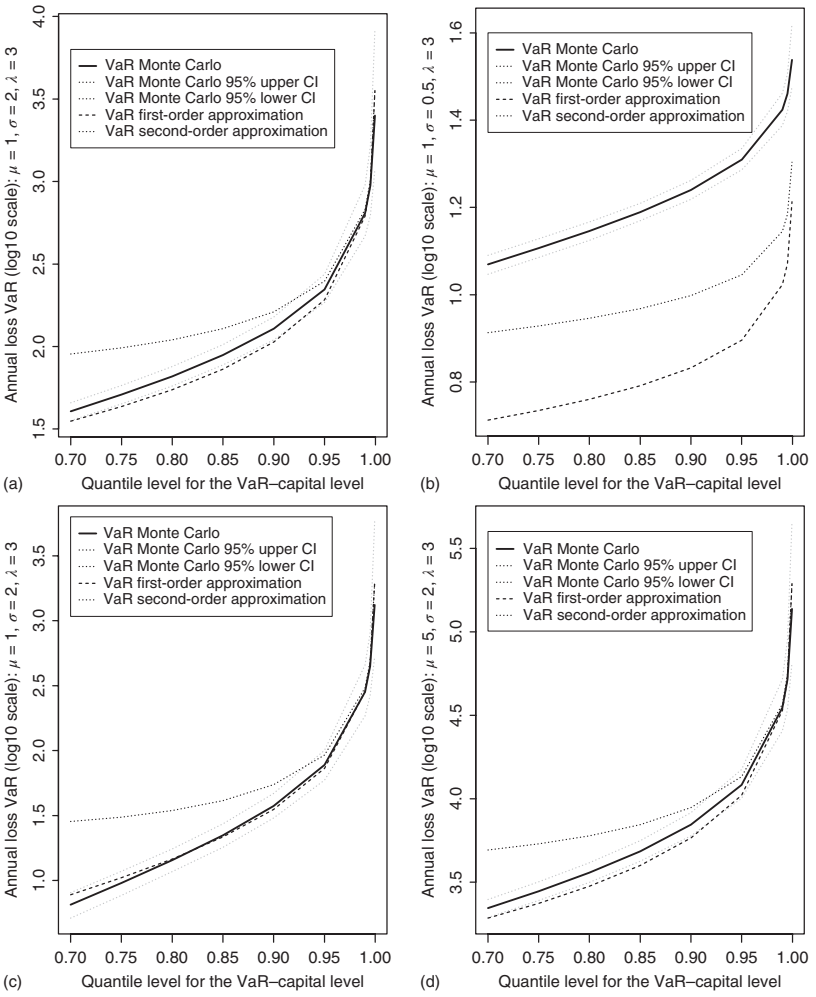

FIGURE 8.2 VaR approximation for the Poisson-LogNormal example.

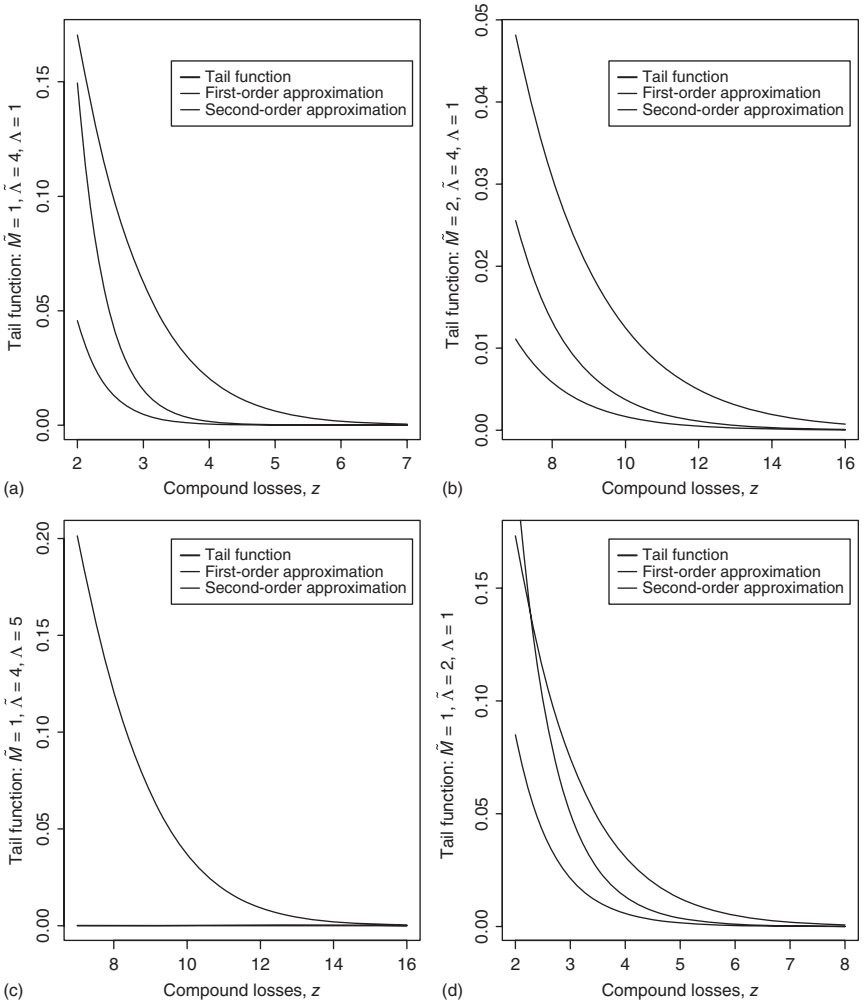

FIGURE 8.3 **Tail function approximation for the Poisson-inverse-Gaussian example.**

Remark 8.15 *For the Poisson-IG case, the greatest sensitivity is clearly on the rate λ (see, e.g. the plot in Figure 8.6c). In difference to the LogNormal example, the second-order approximation seems to perform always better than the first order, but none of them are sufficiently close to any of the 'true' (Monte Carlo) VaR.*

These results, therefore, serve to illustrate the importance of understanding and developing further studies on the rate of convergence of these asymptotic SLAs. This will help to guide in the understanding of when they can be reliably utilized in practice. One can see that even in these relatively simple examples, depending on the values of the parameters in the LDA risk model, the asymptotic VaR approximation may or may not be accurate at quantile levels of interest to risk management. Therefore, even small amounts of parameter uncertainty in the LDA model estimation may manifest in significantly different accuracies in the SLA capital

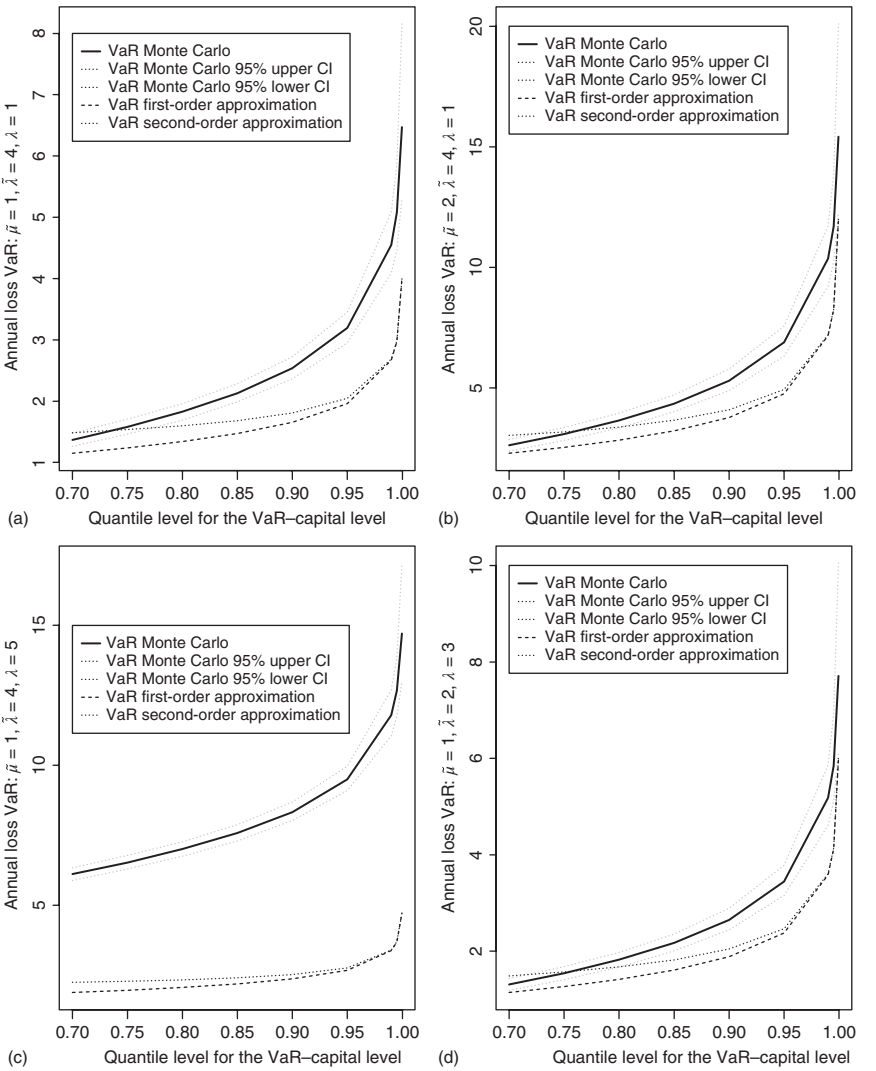

FIGURE 8.4 **VaR approximation for the Poisson-inverse-Gaussian example.**

estimates. The sources of these inaccuracies are discussed in more detail in the conclusion of this chapter where approximation errors are considered.

In addition to these second-order refinements of the SLA-VaR approximation, there has been work developed in Degen and Embrechts (2011) which studies different properties of SLA-VaR refinements and recommends the development of approximations which are improved through the concept of 'penultimate approximations'. In particular, the results developed in this work are focussed on the application of results from EVT to OpRisk VaR estimation. This is discussed in the following subsection as an alternative class of VaR approximations for single risk settings to those presented earlier.

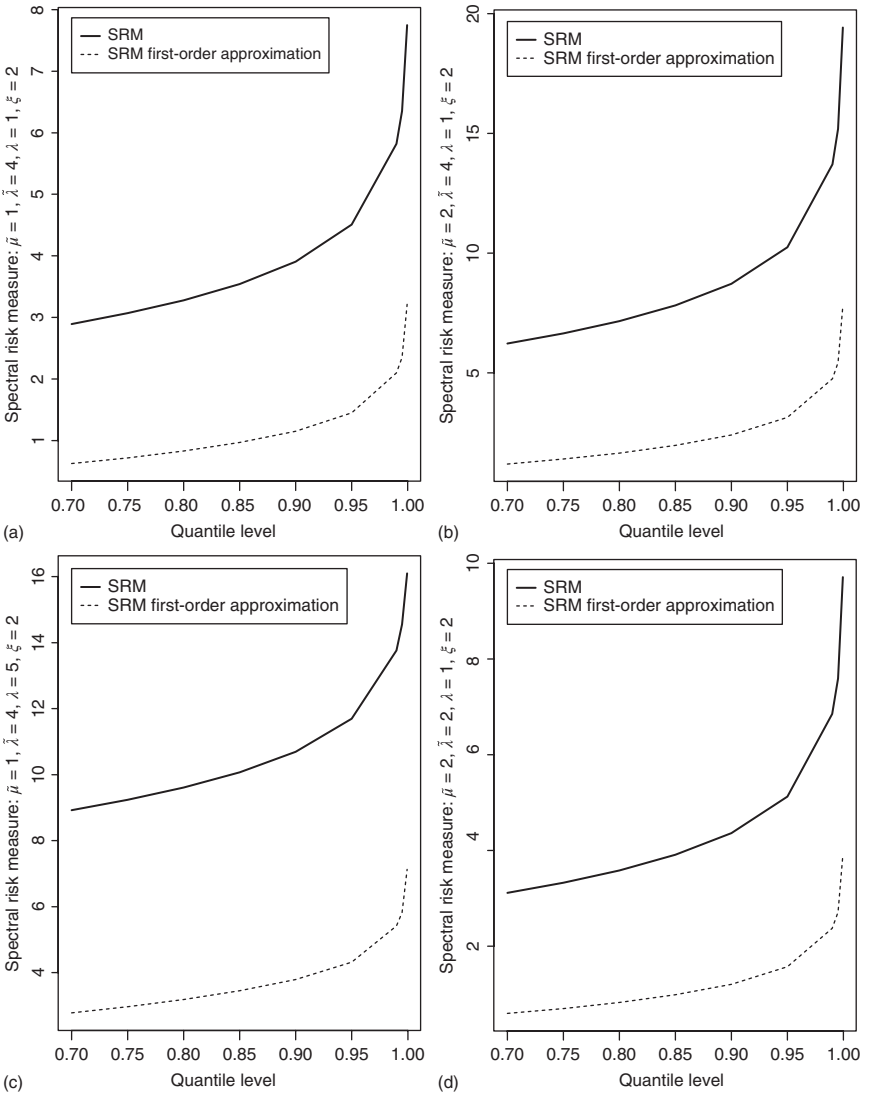

FIGURE 8.5 SRM approximation for the Poisson-inverse-Gaussian example.

8.6 EVT-Based Penultimate SLA

In addition to the refinements made to the SLA-VaR estimates which are based on second-order sub-exponentiality results and regular variation assumptions, there is also a second class of estimators of extreme quantiles that can be utilized in OpRisk settings. The concept of extended regular variation in Definition 3.27 as discussed in Degen and Embrechts (2011) provides an alternative mechanism to develop VaR approximation approaches when utilized in the context of EVT models in a peak-over-threshold (POTs) framework. For required background on EVT models, refer to Chapter 2 and the references therein.

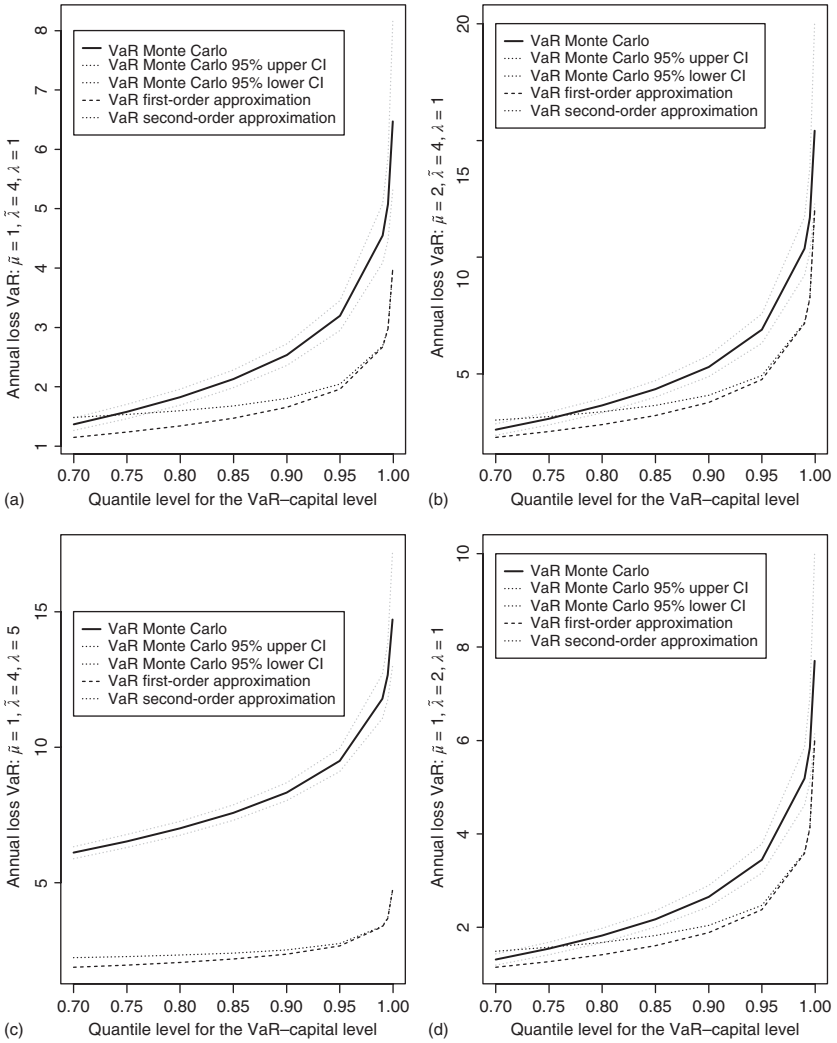

FIGURE 8.6 **VaR approximation for the Poisson-inverse-Gaussian example.**

In particular, the EVT-based estimators discussed next provide alternative SLAs to those developed earlier and are sometimes referred to as penultimate approximations or EVT-based VaR. To begin, one defines the concepts of a linear domain of attraction for the extreme value distribution and a power-normalized domain of attraction as presented in Definition 8.10, see details in Pantcheva (1985), Mohan and Ravi (1993) and Degen and Embrechts (2011, definition 3.1).

Definition 8.10 (Linear and Power-Normalized EVT Distribution Domain of Attraction) *Consider modeling a single risk process via an EVT framework based on a POT method, such that the limiting extreme value distribution* $(H_\gamma(x))$ *of a loss random variable* $X \sim F$ *is obtained in one of the two sequences of scalings for EVI* $\gamma > 0$ *and* $x > 0$*:*

- **Linear Normalization**: *involves the existence of a linear sequence of scaled random variables satisfying the following convergence result:*

$$\mathbb{P}\mathrm{r}\left[\frac{X-t}{f(t)} > x \middle| X > t\right] = \frac{\overline{F}(t+xf(t))}{\overline{F}(x)} \to -\ln H_\gamma(x) = (1+\gamma x)^{-1/\gamma}, \text{ as } t \to x_*.$$

If the severity distribution F satisfies this condition for some measurable function $f(\cdot) > 0$, then the severity distribution belongs to the linear domain of attraction $\mathscr{D}_l^{\mathrm{POT}}$ of the extreme value distribution, denoted by $F \in \mathscr{D}_l^{\mathrm{POT}}(H_\gamma)$.

- **Power Normalization**: *involves the existence of a sequence of power-scaled random variables satisfying the following weak convergence result holds for every point of continuity of x such that*

$$\mathbb{P}\mathrm{r}\left[\left(\frac{X}{t}\right)^{1/g(t)} > x \middle| X > t\right] \to \overline{K}, \text{ as } t \to x_* \tag{8.93}$$

for some non-degenerate distribution K. If the severity distribution F satisfies this condition for some measurable function $g(\cdot) > 0$, then the severity distribution belongs to the power-normalized domain of attraction p^{POT} of K, denoted by $F \in \mathscr{D}_p^{\mathrm{POT}}(K)$. Furthermore, the possible limit laws K are unique up to the p-types such that two distributions K_1 and K_2 of the same p-type satisfy the condition $K_1(x) = K_2(x^p)$ for some $p > 0$.

One can show that there is a one-to-one correspondence between the two domains of attraction under transformation of the loss random variable $X \sim F$ as detailed in Proposition 8.12, see a discussion and detailed proof in Degen and Embrechts (2011, proposition 3.1).

Proposition 8.12 (Equivalent Linear and Power-Normalized Domains of Attraction) *Consider a loss random variable $X \sim F$ and an extreme value index (EVI) $\gamma \in \mathbb{R}$, then the following holds, where $1 - K_\gamma(x) = -\ln H_\gamma(\ln x)$:*

1. *Given a transformed loss random variable $\widetilde{X} = \ln X$ with a distribution $\widetilde{X} \sim F_{\widetilde{X}} \in \mathscr{D}_l^{\mathrm{POT}}(H_\gamma)$, it is equivalent to state the following membership of the untransformed loss random variable in the power-normalized domain of attraction such that $F_X \in \mathscr{D}_p^{\mathrm{POT}}(K_\gamma)$.*
2. *Secondly, if $X \sim F_X \in \mathscr{D}_l^{\mathrm{POT}}(H_\gamma)$, then it is equivalent to state the $X \sim F_X \in \mathscr{D}_p^{\mathrm{POT}}(K_{\min(\gamma,0)})$.*

These relationships are expressed in terms of the quantile functions via application of Resnick (2007, proposition 0.1). These provide the equivalences for tail quantile functions $U_X(x) = F^{\leftarrow}(1-1/x)$ given for an auxiliary measurable function $a(\cdot) > 0$ by the following memberships of the class of Extended Regularly Variation (ERV) functions:

1. $U_{\widetilde{X}} \in \mathrm{ERV}_\gamma(a) \Longleftrightarrow \ln U_X \in \mathrm{ERV}_\gamma(a)$, *and*
2. $U_X \in \mathrm{ERV}_\gamma(a) \Rightarrow \ln U_X \in \mathrm{ERV}_{\min(\gamma,0)}\left(\frac{a(t)}{U(t)}\right)$.

The definition of the family of Extended Regular Variation (ERV) functions is provided in Section 3.5.

With these definitions and equivalences, we start with a simple EVT result for VaR and ES capital estimation based on results from Marinelli *et al.* (2007). As discussed in the section on EVT models in Chapters 2, there are methods to estimate the VaR and ES based on only the assumption that the loss process has loss random variables which come from a distribution which is in the domain of attraction of a max-stable law.

Proposition 8.13 (Estimated Tail and VaR via Hill Estimators and EVT Asymptotics) *Consider i.i.d. losses from a compound process that are ordered to give the set of n-loss random variable order statistics $\{X_{(i,n)}\}_{i=1}^{n}$. Using these order statistics forms an estimate of the EVI denoted by $\widehat{\gamma}$ via, for instance, the Hill estimator according to*

$$\widehat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \ln X_{(i,n)} - \ln X_{(k,n)}, \tag{8.94}$$

where $X_{(n,n)} \leq X_{(n-1,n)} \leq \cdots \leq X_{(1,n)}$ and some selected index k is obtained see discussion in chapter 2. Then an estimate of the tail distribution is obtained via

$$\widehat{\overline{F}}_Z(x) = \widehat{\mathbb{P}\mathrm{r}}[Z > x] = \frac{k}{n}\left(\frac{x}{X_{(k+1,n)}}\right)^{-\widehat{\gamma}^{-1}} \tag{8.95}$$

and the estimation of the VaR or quantile (denoted by x_p) for probability $p \in (0,1)$ can be obtained via several different estimators each based on the estimated EVI parameter according to one of the following expressions:

- *Marinelli* et al. *(2007) considered the estimator given by*

$$\widehat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-\widehat{\gamma}} X_{(k+1,n)}. \tag{8.96}$$

- *Dekkers and De Haan (1989) and Embrechts* et al. *(1997) considered the estimator given by*

$$\widehat{x}_p = X_{(k,n)} + (X_{(k,n)} - X_{(2k,n)})\frac{\left(\frac{k}{n(1-p)}\right)^{-\widehat{\gamma}} - 1}{1 - 2^{-\widehat{\gamma}}}. \tag{8.97}$$

For the VaR or quantile estimator in Proposition 8.13 proposed in Marinelli *et al.* (2007), one can also state the following results given in Theorem 8.4 which provides a convergence in distribution for the translated and scaled VaR estimators distribution to normality. This result provides the ability to obtain an asymptotic confidence interval for the estimator.

Theorem 8.4 (Confidence Interval for VaR Estimator based on EVT Asymptotics) *Consider a loss process with n losses which are i.i.d. such that $X_i \sim F(x)$ where the population distribution $F(x)$ satisfies that it is in the domain of attraction of a generalized extreme value (GEV) distribution $F(x) \in \mathcal{D}(H_\gamma)$ for a strictly positive EVI parameter $\gamma > 0$. In addition, assume that the population distribution for the loss process F admits a positive density f which satisfies $f \in RV_{-1-1/\gamma}$. Assuming the convergence $n(1-p) \to c$ to a fixed constant $c > 0$ is satisfied and using the EVI estimator given by the Pickands estimator $\widehat{\gamma}$ (see Chapter 2), then for every constant $k > c$, one has convergence in distribution to random variable ν given by*

$$\frac{\widehat{x}_p - x_p}{X_{(k,n)} - X_{(2k,n)}} \Rightarrow \nu, \tag{8.98}$$

where ν is given by

$$\nu = \frac{\left(\frac{k}{c}\right)^{\gamma} - 2^{-\gamma}}{1 - 2^{-\gamma}} + \frac{1 - \left(\frac{Q_k}{c}\right)^{\gamma}}{\exp(\gamma H_k) - 1}, \tag{8.99}$$

for independent random variables $Q_k \sim \exp(2k+1)$ *and* $H_k = \sum_{i=k+1}^{2k} \frac{1}{i} E_i$ *with i.i.d.* $E_i \sim Exp(1)$ *for* $i = 1, 2, \ldots$

In addition to these basic EVT-based VaR estimators and the confidence interval, one can also develop the following second-order asymptotics of normalized quantiles of the single-loss POT model may be proven, see Gomes and De Haan (1999) and Degen and Embrechts (2011).

Lemma 8.1 (EVT Quantile Approximations for Extended Regular Variation) *Given a quantile function* $U_X(x) = F^{\leftarrow}(1 - 1/x)$ *with* $U \in ERV_\gamma(a)$ *for* $\gamma \geq 0$ *and* $x > 0$*, then the following two different quantile function approximations can be obtained through considering Proposition 8.12 for* U *and alternatively for* $\ln U$ *as follows:*

1. **Quantile Function Approximation 1**. *The condition* $U \in ERV_\gamma(a)$ *implies by definition that*

$$\frac{U(tx) - U(t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}, \; t \to \infty, \tag{8.100}$$

which gives the approximation for the tail quantile function given by

$$U(tx) \approx U(t) + a(t)\frac{x^\gamma - 1}{\gamma}. \tag{8.101}$$

2. **Quantile Function Approximation 2**. *The condition* $U \in ERV_\gamma(a)$ *implies* $\ln U \in ERV_{\min(\gamma,0)}\left(\frac{a(t)}{U(t)}\right)$ *as detailed in Proposition 8.12, which, therefore, gives for* $\frac{a(t)}{U(t)} > 0$ *the convergence*

$$\left(\frac{U(tx)}{U(t)}\right)^{\frac{U(t)}{a(t)}} \to x, \; t \to \infty, \tag{8.102}$$

when $\frac{a(t)}{U(t)} \to \gamma$*. This gives the approximation for the tail quantile function given by*

$$U(tx) \approx x^{\frac{a(t)}{U(t)}} U(t). \tag{8.103}$$

In the risk management context, this implies the two following VaR-EVT penultimate approximations given in Theorem 8.5 which are based on the scaling properties of high quantile tail function estimators. It is important to note that in order to avoid complications with the following penultimate approximations arising from the lack of smoothness of the quantile function, in Degen and Embrechts (2011), a set of sufficient conditions are proposed for the suitable application of the EVT penultimate approximations, as presented in Proposition 8.14.

Proposition 8.14 (Sufficient Conditions on Quantile EVT Penultimate Approximations) *Consider a single-loss process in which the POTs EVT-based modeling approach is considered for severity distribution* $X \sim F(x)$ *which has a tail quantile function given by the generalized inverse* $U(t) = F^{\leftarrow}(1 - 1/t)$ *which satisfies the following conditions:*

- *$U(t)$ can be represented by $U(t) = \exp(\phi(\ln t))$ where*

$$\phi(t) = \int_1^{\exp(t)} \frac{ds}{u(s)} + c, \tag{8.104}$$

with $u(s) = \frac{U(s)}{U'(s)}$ and $c = \ln U(1)$.

- *von Mises condition holds where for some $\gamma \geq 0$, the following convergence asymptotically holds*

$$\frac{tU''(t)}{U'(t)} \to \gamma - 1, \tag{8.105}$$

which ensures the log–log plot of ϕ for tail quantile function U is linear with converging gradient given by ϕ' and diminishing convexity $\Phi'' \to 0$ asymptotically.

If these conditions are satisfied (see sufficient conditions in Degen and Embrechts (2011, proposition 4.2)), then one may safely consider the following POTs-based penultimate SLAs.

Theorem 8.5 (Penultimate Singe-Loss Approximate VaR for POTs Models) *Given a loss random variable $X \sim F(x)$ such that the quantile function satisfies $U \in ERV_\gamma(a)$ for $\gamma \geq 0$ and $x > 0$, then the POT-EVT annual loss model standard VaR is estimated for a measurable function f and a threshold $u = VaR_{\widetilde{\alpha}}^{POTI}$ by*

$$\mathrm{VaR}_\alpha^{\mathrm{POTI}} \approx u + \widehat{f}(u)\frac{\left(\frac{N_u}{n(1-\alpha)}\right)^{\widehat{\gamma}} - 1}{\widehat{\gamma}}, \text{ with } \widetilde{\alpha} < \alpha, \tag{8.106}$$

and estimates $\widehat{\gamma}$ and $\widehat{f}(u)$ for EVI γ and POT-EVT scale parameter $f(u)$, where N_u is the number of loss exceedances above threshold u. Furthermore, $\widehat{\overline{F}(u)}$ is estimated by N_u/n where n is the total number of losses. In addition, when $U \in ERV_\gamma(a)$, then the implied condition that $\ln U \in ERV_{\min(\gamma,0)}\left(\frac{a(t)}{U(t)}\right)$ gives rise to the penultimate EVT-POT VaR approximation given by

$$\mathrm{VaR}_\alpha^{\mathrm{POTII}} \approx x^{\widehat{b(t)}}u, \tag{8.107}$$

where $b(t) = \frac{a(t)}{U(t)}$ is estimated as proposed in Degen and Embrechts (2011) by either of the following two choices:

- *Estimator 1: $b(t) = \phi'(\ln t)$ for quantile level α;*
- *Estimator 2: $b(t) = \widetilde{\phi}'(\ln t)$ for quantile levels $\widetilde{\alpha} < \alpha$;*

with $t = \frac{1}{1-\widetilde{\alpha}}$ and $x = \frac{1-\widetilde{\alpha}}{1-\alpha} > 1$.

Alternative asymptotic approximations of the VaR for Poisson-Pareto models have also been developed in the literature, see, for example, Böcker and Klüppelberg (2005). Under these approaches, it is proposed that in cases in which sufficient data is available and accurate, estimation techniques discussed in Chapter 2 are utilized to estimate the EVI; then one can also utilize such results to obtain an accurate single-loss-type approximation of the VaR.

Theorem 8.6 (Point Process VaR Estimate) *Consider single risk LDA OpRisk model with severity distribution $F_X(x)$ which is of slow variation and can be expressed according to*

$\overline{F}_X(x) \sim \frac{x}{\theta}^{-\alpha} L(x)$ as $x \to \infty$ for some slowly varying function $L(x)$. Furthermore, assume a number N_t of i.i.d. losses X_i in a time period $[0,t]$ are observed with the maximum denoted by the order statistic $X_{(N_t)}$. Then a point-process-based estimator of the VaR for an LDA OpRisk model constructed from such a Poisson-severity model combination will be obtained for a level κ quantile of the annual loss by

$$\mathrm{VaR}_t[\kappa] \approx \left(\frac{1+\frac{1}{\alpha}}{1-\kappa}\right)^{\frac{1}{\alpha}} x_m, \tag{8.108}$$

where $x_m = \mathrm{mode}(Z_{(N_t)})$ and $Z_{(N_t)}$ denotes the maximum annual loss in $[0,t]$ and x_m is obtained as the solution to

$$\lambda t \exp(-\lambda t \overline{F}_X(x)) \left[\lambda t f_X^2(x) + \frac{d^2}{dx^2} F_X(x)\right] = 0. \tag{8.109}$$

Remark 8.16 *It should be noted that the VaR approximation obtained is not a function of the frequency process, but critically depends on two variables: the ability to accurately estimate the tail index α and the confidence interval for which approximation is performed for the annual loss quantile. Note that the accurate estimate of α could require a significant amount of loss data or, alternatively, a reasonably informed prior distribution in a Bayesian context.*

To understand where this result is obtained, we step through the process utilized to arrive at this result. If one considers a quantile of the annual loss distribution at a level α, that is, the potential annual loss that is exceeded only with small probability $1-\alpha$. Assuming a number N_t of i.i.d. losses X_i in a time period $[0,t]$ are observed, then the maximum denoted by the order statistic $X_{(N_t)}$ can be used to calculate the distribution of the maximum of the annual loss LDA compound process model according to

$$\begin{aligned} F_{Z_{(N_t)}}(x) =& \mathbb{P}\mathrm{r}[Z_{(N_t)} \le x] \\ =& \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N_t = n] \mathbb{P}\mathrm{r}[Z_{(n)} \le x] \\ =& \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N_t = n] F_X^n(x), x \ge 0, \end{aligned} \tag{8.110}$$

for the single risk LDA models with a severity model which can be assumed to be of slow variation, in which $\overline{F}_X(x) \sim (x/\theta)^{-1/\gamma}$ as $x \to \infty$. Severity models of this type include the GPD severity models motivated in this chapter and combines these severity models under the Poisson-Pareto-type LDA model; then the distribution of the annual maximum loss is given by

$$\begin{aligned} F_{Z_{(N_t)}}(x) =& \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}(N_t = n) F_X^n(x), x \ge 0 \\ =& \exp(-\lambda t) \sum_{n=0}^{\infty} \frac{(\lambda t F_X(x))^n}{n!} \\ =& \exp(-\lambda t (1 - F_X(x))) \\ =& \exp(-\lambda t \overline{F}_X(x)),\ x \ge 0. \end{aligned} \tag{8.111}$$

In this model for the distribution of the maximum loss up to time t, given by $F_{Z_{(N_t)}}(x)$, the most probable value $x_m = \text{mode}(Z_{(N_t)})$. If the derivative of the severity models density function exists, then the derivative of the density of the maximum loss $f_{Z_{(N_t)}}(x)$ also exists and is given by

$$\frac{d^2}{dx^2} F_{Z_{(N_t)}}(x) = \lambda t \exp(-\lambda t \overline{F}_X(x)) \left[\lambda t f_X^2(x) + \frac{d^2}{dx^2} F_X(x) \right]. \tag{8.112}$$

Hence, the mode (most probable annual loss value x_m) is given by equating this to zero and numerically solving the non-linear equation via a root search method such as the Newton or bisection approach for the equation

$$\lambda t f_X^2(x) + \frac{d^2}{dx^2} F_X(x) = 0. \tag{8.113}$$

In many severity models, the solution to this equation will produce a unique result as most severity models in OpRisk are typically univariate. For example, the case of the Poisson-generalized Pareto-LDA model discussed in this chapter one would obtain a solution for x_m given by

$$x_m = \theta \left[\left(\frac{\alpha \lambda t}{1+\alpha} \right)^{\frac{1}{\alpha}} - 1 \right] \approx \theta \left(\frac{\alpha \lambda t}{1+\alpha} \right)^{\frac{1}{\alpha}}. \tag{8.114}$$

Hence, the OpRisk VaR is then approximated, in terms of the EVI tail index, as

$$\text{VaR}_t[\kappa] \approx \left(\frac{1+\frac{1}{\alpha}}{1-\kappa} \right)^{\frac{1}{\alpha}} x_m = \theta \left(\frac{1+\frac{1}{\alpha}}{1-\kappa} \right)^{\frac{1}{\alpha}} \left(\frac{\alpha \lambda t}{1+\alpha} \right)^{\frac{1}{\alpha}}. \tag{8.115}$$

Next we move to the approximation of two other risk measures, the ES and the more general SRM. The reasons alternative risk measures are being proposed to replace the VaR as regulatory mandated calculation for capital is primarily due to the fact that essential disadvantage of the VaR risk measure is that, in general, it is not coherent. Therefore, situations may arise in which VaR attributes more risk to a loss portfolio than to the sum of the single-loss positions; this is what is meant by incoherency of risk measures, as discussed in Cruz *et al.* (2015, chapter 6). In addition, one must also consider the fact that VaR exclusively regards the probability of a loss in measuring the risk or capital required, whereas the size or magnitude of the losses remains out of consideration.

8.7 Motivation for Expected Shortfall and Spectral Risk Measures

Over recent years in OpRisk and other areas of risk management, there has been a shift from just considering VaR risk measures to also considering alternative measures of risk that can be used to quantify OpRisk capital. As noted in Cruz *et al.* (2015, chapter 6), the VaR is not a sub-additive risk measure (i.e. it is not coherent). It was noted in Acerbi (2004) that

> "The main problem with VaR is not its lack of subadditivity, but rather the very fact that no set of axioms for a risk measure and therefore no unambiguous definition of financial risk has ever been associated with this statistic. So, despite the fact that some VaR supporters still claim that subadditivity is not a necessary axiom, none of them, to the best of our knowledge, has ever tried to write an alternative meaningful and consistent set of axioms for a risk measure which are fulfilled also by VaR."

As a consequence of this realization, several authors explored alternative measures of risk to VaR which aimed to provide alternative risk measures that retained the benefits of VaR with respect to its global application to any loss process and its probabilistic grounding. As noted in the introduction to this chapter, one of the possible alternative candidate risk measures that have taken prominence after VaR (although it is not a regulatory standard) is the ES. As was provided in its definition previously in Equation 8.2, it represents the average of the worst losses.

The ES has many names in the actuarial and risk management literature: expected tail loss, tail conditional expectation, conditional VaR, tail conditional VaR and worst conditional expectation. It is worth noting that in the case that the loss distribution is continuous, then all such measures will coincide with the ES; however, when discrete distributions or atoms are present, there may be differences in each of these definitions, and in this case, the ES is known to always be unique and coherent, where as the other definitions may differ in these properties, see discussions on the technicalities of these different versions of risk measure in Artzner *et al.* (1997), Artzner *et al.* (1999) and Pflug (2000).

In the context of OpRisk modeling, the ES has been studied for such data in Biagini and Ulmer (2009) where they note that

> "The risk measure Expected Shortfall is more conservative than the VaR at the same confidence level. Since it can be assumed that the actual loss profile exhibits several extremely high losses with a very low probability, the Expected Shortfall is the more appropriate risk measure, as, in contrast to the VaR, it regards the size of this extreme losses."

As noted in Biagini and Ulmer (2009), the ES appears as the best convex alternative risk measure to the VaR. The notion of best here is meant in the sense that ES will be the smallest law-invariant, convex risk measure which is continuous from below and dominates VaR. It is also worth noting that under the Solvency II insurance regulations, the ES is required to be evaluated under the Swiss Solvency test.

In Dowd and Blake (2006), they also motivate the use of ES because it can be argued to form a justification for scenario analysis and stress testing which is critical to many areas of OpRisk modeling where rare high consequence loss data is scarce and hard to come by. They argue that the results of scenario analyses may be interpreted as coherent risk measures. To understand this, consider the set of losses generated under a scenario analysis framework and the associated probabilities of such events. For a given loss threshold, such as a VaR level, the loss events in the scenario analysis that exceed the specified VaR level are then drawn from the tail of the unspecified loss distribution. Hence, the average of these losses can be considered as estimators for the ES associated with the loss distribution. The interesting point of this is that because ES is a coherent risk measure, then this implies that outcomes of scenario analyses are also coherent risk measures.

As noted in Dowd and Blake (2006), one may also show that the class of coherent risk measures contains the ES and associated risk measures given by

1. the highest expected loss from a set of possible outcomes such as from a scenario analysis;
2. the highest ES from a set of comparable ESs which are each based on different distribution function or LDA model; and
3. the highest expected loss from a set of highest losses.

It should, however, be pointed out that given an LDA model structure with heavy tails, the accurate estimation of the ES can be computationally expensive because of the large number of simulations required to obtain improbable high consequence loss realizations that will be utilized in the conditional tail expectation. Hence, it is practically very useful to have access to closed-form representations such as the SLA results presented in the following section.

In Tong and Wu (2012), it was observed that the VaR and the ES risk measures are unable to make allowances for international banks to include their own risk attitudes or appetites. They argue that the SRM developed in Acerbi (2002) is a good alternative because it allows for banks to incorporate such risk appetite quantifications into the capital estimation. In addition, the SRM is also a coherent risk measure and contains as a special case the class of ES risk measures.

Dowd and Blake (2006) comment that when considering the class of SRM, one aims to employ a risk measure framework which also combines properties of risk-aversion theory. This requires that the OpRisk modeller would be willing to specify quantitatively as a function their risk aversion. In some contexts, this may be inappropriate or controversial because it may produce problems such as how does a company or financial institution hope to quantify a homogeneous and consistently agreed on risk-aversion function amongst all business units with different functions and different risk appetites. This could involve disagreements and ambiguity regarding the most appropriate form for the risk-aversion function, and in addition, even when a risk-aversion function is agreed upon, what are the parameter values for this function that are appropriate and how should one elicit them to remain conservative in the resulting capital measure.

Recall that the SRM has the form

$$\mathrm{SRM}_{\phi_\kappa}[Z] = \int_\kappa^1 \phi_\kappa(s)\mathrm{VaR}_s[Z]\,ds, \tag{8.116}$$

for weighting function (risk spectrum or risk-aversion function) $\phi_\kappa(s)$. In Dowd and Blake (2006), they explain that the following three properties are required to be satisfied in order for SRM to be coherent:

1. *Non-negativity*. The risk-aversion function $\phi_\kappa(s) \geq 0$ for all $\kappa, s \in [0,1]$.
2. *Normalization*. The risk-aversion function $\phi_\kappa(s)$ should be normalized as follows:

$$\int_0^1 \phi_0(s)\,ds = 1. \tag{8.117}$$

3. *Increasing*. The risk-aversion function $\phi_\kappa(s)$ should be increasing such that for any $\kappa \in (0,1)$ one has $\phi_\kappa(s_1) \leq \phi_\kappa(s_2)$ for all $\kappa \leq s_1 \leq s_2 \leq 1$.

These should, therefore, act as a minimal set of requirements for risk managers to consider when specifying their risk-aversion function. The last condition simply implies that larger losses should be no smaller than weights attached to smaller loss amounts. This last point, simple as it may be, is the key to coherency of ES and SRM and also the reason why VaR fails to be a coherent risk measure.

Before completing this brief motivation for the consideration of ES and SRM in OpRisk settings as alternatives to VaR, it is also worthwhile making clear the connection between ES and the expectile risk measures discussed in Newey and Powell (1987) and in the context of risk measures in the Conditional Autoregressive Expectile (CARE) framework of Taylor (2008) where it is explained how to perform semi-parametric estimation of the VaR and ES via the link to expectiles which are specified in Definition 8.11, see Aigner *et al.* (1976).

Definition 8.11 (Expectiles) *The κ-level expectile for a random variable Z, denoted by μ_κ, is a parameter that minimizes the expectation given by*

$$\mathbb{E}[(Z-\mu_\kappa)^2|\kappa - \mathbb{I}[Z<\mu_\kappa]|], \tag{8.118}$$

for $\kappa \in (0,1)$.

Remark 8.17 *Generally, the κ-level expectile μ_κ is neither the VaR nor the ES and does not have a simple intuitive explanation.*

One can observe that μ_κ occurs at a quantile level of the annual loss Z, denoted by q_κ, and typically one has that $\kappa < q_\kappa$, and hence, one can also see that μ_κ also minimizes the following expectation

$$\mathbb{E}[(q_\kappa - \mathbb{I}[Z<\mu_\kappa])(Z-\mu_\kappa)]. \tag{8.119}$$

It was then shown in Newey and Powell (1987) that there is a one-to-one relationship between the expectiles and the ES risk measure. In the simple case that $\mathbb{E}[Z]=0$, one would obtain the relationship

$$\mathrm{ES}_\alpha[Z] = \left(1+\frac{\kappa}{(1-2\kappa)q_\kappa}\right)\mu_\kappa. \tag{8.120}$$

Note that typically, this will not be a case of interest in OpRisk settings, and more general expressions may be obtained in above-mentioned reference. This relationship provides interesting alternative statistical methods to perform estimation of ES for OpRisk settings based on quantile regressions.

8.8 First- and Second-Order Approximation of Expected Shortfall and Spectral Risk Measure

In this section, we consider developments in SLAs of first order and second order that have been proposed in the literature for alternative capital risk measures given by the SRM and the special sub-case of the ES. As discussed in the motivation to this section, ES is an important quantity to also consider computation of for LDA models as it is well known to be a convex and coherent risk measure, unlike the VaR risk measure, see discussions in Cruz *et al.* (2015, chapter 6). Hence, we begin this discussion by considering the class of risk measures known as ES and then move to the class of SRM which we briefly recall were defined by

1. *Spectral Risk Measure (SRM).* For a weight function $\phi:[0,1]\mapsto\mathbb{R}$, SRM is given by

$$\mathrm{SRM}_\phi[Z] = \int_0^1 \phi(s)\mathrm{VaR}_s[Z]ds, \tag{8.121}$$

for suitable families of function $\phi(x)$.

2. *Expected Shortfall (ES).* For a level $\alpha \in (0,1)$, if there is no jump in distribution at $\mathrm{VaR}_\alpha[Z]$, the ES is given by the tail conditional expectation of the annual loss distribution according to

$$\mathrm{ES}_\alpha[Z] = \mathbb{E}[Z|Z \geq \mathrm{VaR}_\alpha[Z]] = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s[Z]\,ds. \tag{8.122}$$

We will start with the more widely considered class of risk measures the ES and proceed with the SRM class, making clear along the way how the asymptotics are derived and how they relate to the first- and second-order asymptotics derived earlier for the annual loss VaR. Before presenting the results on the asymptotic approximations of these risk measures, it is worthwhile to recall a few basic definitions that were presented in previous chapters but are recalled here for convenience.

Proposition 8.15 (Asymptotic Equivalence of Summands and Summations) *Given two functions (distributions) $F_i(x)$ and $G_i(x)$ which are asymptotically equivalent $F_i(x) \sim G_i(x)$ as $x \to \infty$ such that*

$$\lim_{x\to\infty} \frac{F_i(x)}{G_i(x)} = 1, \tag{8.123}$$

which is equivalent to saying $F_i(x) = G_i(x)(1 + o_i(1))$ as $x \to \infty$. If one then considers $i = 1, 2, \ldots, n$ of these functions, then the asymptotic equality of the summands provides the asymptotic equality of their sums. Therefore, one has under summation the asymptotic equivalence of the sums when the individual summands are each asymptotically equivalent for all $i = 1, 2, \ldots, n$ written in either of the following equivalent forms:

$$F_1(x) + F_2(x) + \cdots + F_n(x) \sim G_1(x) + G_2(x) + \cdots + G_n(x), \text{ as } x \to \infty, \tag{8.124}$$

or

$$\lim_{x\to\infty} \frac{F_1(x) + F_2(x) + \cdots + F_n(x)}{G_1(x) + G_2(x) + \cdots + G_n(x)}$$

$$= 1 + \frac{G_1(x)o_1(1) + G_2(x)o_2(1) + \cdots + G_n(x)o_n(1)}{G_1(x) + G_2(x) + \cdots + G_n(x)} = 1. \tag{8.125}$$

One can also state a similar useful result for integrands and integrals as follow in Proposition 8.16.

Proposition 8.16 (Asymptotic Equivalence of Integrands and Integrals) *Consider two functions $\varphi(\alpha)$ and $\psi(\alpha)$, each mapping $[0,1] \mapsto [0,\infty)$, which are asymptotically equivalent $\varphi(\alpha) \sim \psi(\alpha)$ as $\alpha \to 1$ such that*

$$\lim_{\alpha\uparrow 1} \frac{\varphi(\alpha)}{\psi(\alpha)} = 1. \tag{8.126}$$

Furthermore, assume that there exists a constant $\tau \in [0,1)$ such that the functions each satisfy the integrals

$$\begin{aligned} \int_\tau^1 \varphi(s)\,ds < \infty, \\ \int_\tau^1 \psi(s)\,ds < \infty, \end{aligned} \tag{8.127}$$

then one has the asymptotic equivalence of the following integrals:

$$\int_{\alpha}^{1} \varphi(s)\,ds \sim \int_{\alpha}^{1} \psi(s)\,ds, \tag{8.128}$$

as $\alpha \uparrow 1$.

The following Proposition 8.2 allows one to link the asymptotic equivalence between two tails of distribution functions to the asymptotic equivalence of integrals formed with respect to these distributions, see Biagini and Ulmer (2009, lemma 2.6).

Lemma 8.2 (Asymptotic Equivalence of Tail Distributions and their Integrals) *Consider two distribution functions* F *and* G *that have tail functions which are asymptotically equivalent up to a scalar given by*

$$\overline{F}(x) \sim C\overline{G}(x), \tag{8.129}$$

with $x \to \infty$ *for some constant* $C \in \mathbb{R} \notin \{0\}$. *Then for every measurable function* $\psi(x)$ *that maps* $\mathbb{R} \mapsto \mathbb{R}$ *and is integrable with respect to distributions* F *and* G, *the following applies:*

$$\int_{q}^{\infty} \psi(x)\,dG(x) \sim C \int_{q}^{\infty} \psi(x)\,dF(x), \tag{8.130}$$

as $q \to \infty$.

It will also be relevant to recall the following asymptotic equivalence statement between two regularly varying functions and their integrals given in Proposition 8.17, see Biagini and Ulmer (2009).

Proposition 8.17 (Asymptotic Equivalence of RV Functions and Integrals) *Consider two regularly varying functions* $U \in RV_{\rho}$ *and* $V \in RV_{\rho}$ *for some constant of regular variation* $\rho < -1$. *Furthermore, assume that the two functions are asymptotically equivalent where* $U(x) \sim V(x)$ *as* $x \to \infty$. *In this case, one also has the asymptotic equivalence of the integrals of these functions given by the following expression:*

$$\int_{q}^{\infty} U(x)\,dx \sim \int_{q}^{\infty} V(x)\,dx, \quad q \to \infty. \tag{8.131}$$

Finally, we also recall the property of asymptotically equivalent functions which are then embedded as composite functions in a class of regularly varying functions and the resulting asymptotic equivalence given in Proposition 8.18.

Proposition 8.18 (Asymptotic Equivalence of RV Functions and Composite Functions) *Consider two positive functions on support* $(0, \infty)$ *that are asymptotically equivalent* $f(x) \sim cg(x)$ *as* $x \to \infty$ *with each function* $f(x) \to \infty$ *and* $g(x) \to \infty$ *as* $x \to \infty$. *Given a function* U *which is regularly varying* $U \in RV_{\rho}$ *for* $\rho \in \mathbb{R}$, *one can show the following asymptotic equivalence of the composite functions*

$$U(f(x)) \sim c^{\rho} U(g(x)), \quad x \to \infty. \tag{8.132}$$

Using these asymptotic equivalence results and the following monotone density Theorem 8.7 (see Bingham *et al.* 1989), one may derive a closed-form expression for the ES of an LDA model if the severity distribution satisfies that it is in the class of regularly varying functions.

Theorem 8.7 (Monotone Density Theorem) *Consider a function $U(x)$ defined by the integral relationship*

$$U(x) = \int_x^\infty u(s)ds, \tag{8.133}$$

for a function u that is ultimately monotone such that for some $x > 0$, the function $u(x)$ is monotone on (x, ∞). If the function $U(x)$ is regularly varying such that it is asymptotically equivalent to the slowly varying function L giving

$$U(x) \sim cx^\rho L(x), \quad x \to \infty, \tag{8.134}$$

then if $c \in \mathbb{R}$, $\rho \in \mathbb{R}$ and $L \in RV_0$, one has the following asymptotic equivalence for the ultimately monotone function $u(x)$ (the integrand) given by

$$u(x) \sim -c\rho x^{\rho-1} L(x), \quad x \to \infty. \tag{8.135}$$

8.8.1 UNDERSTANDING THE FIRST-ORDER SLA FOR ES FOR REGULARLY VARYING LOSS MODELS

To derive the first-order closed-form asymptotic approximation for the ES, the approach of Biagini and Ulmer (2009) was to assume that the LDA model has a severity which is in the class of heavy-tailed loss models, in particular with right tail that is in the class of regularly varying functions. The additional results they utilized to make this derivation involved properties of regular varying functions that were discussed in Chapter 3. In the following key steps, we outline the basic components utilized to derive the ES first-order SLA, see further discussion in Biagini and Ulmer (2009).

1. Consider a severity distribution F_X which admits a Lebesgue density that is continuous. The continuity of F will also lead to the result that $F_X(F_X^{\leftarrow}(x)) = x$ for $x > 0$. In addition, it will imply that annual loss distribution given by

$$F_Z(x) = \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N = n] F_X^{(n)*}(x) \tag{8.136}$$

with $F_X^{(0)*} = \mathbb{I}_{[0,\infty)}$will also be continuous because of monotone convergence of sums. Under these assumptions, given this annual loss distribution, one can then write the ES as follows:

$$\begin{aligned} \mathrm{ES}_\kappa[Z] &= \mathbb{E}[Z | Z > \mathrm{VaR}_\kappa[Z]] \\ &= \frac{\mathbb{E}[Z\mathbb{I}_{Z>q_\kappa}]}{\mathbb{P}\mathrm{r}[Z > q_\kappa]} \\ &= \frac{1}{1-\kappa} \int_{q_\kappa}^\infty x dF_Z(x), \end{aligned} \tag{8.137}$$

with $q_\kappa := \mathrm{VaR}_\kappa[Z]$.

2. Assume the following properties of the frequency and severity distributions in the LDA model for which the ES is being calculated.

- The severity distribution F_X is sub-exponential and in the class of regularly varying functions such that $\overline{F}_X \in RV_{-\alpha}$ for some constant $\alpha > 1$. This will imply that it will admit a representation given by

$$\overline{F}_X(x) = x^{-\alpha} L(x), \quad x > 0, \tag{8.138}$$

for some slowly varying function $L \in RV_0$. For such a severity model assumption, via the monotone density theorem and setting $U = \overline{F}_X$, $\rho = -\alpha$ and $c = 1$, one has the following asymptotic equivalence for the severity density:

$$f_X(x) \sim \alpha x^{-\alpha-1} L(x), \quad x \to \infty. \tag{8.139}$$

- The frequency distribution in the LDA model satisfies the condition that

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n \mathbb{P}\mathrm{r}[N = n] < \infty. \tag{8.140}$$

3. For such assumptions on the LDA model components, we have from Chapter 7 that the annual loss distribution tail is asymptotically equivalent to the tail of the severity

$$\overline{F}_Z(x) \sim \mathbb{E}[N] \overline{F}_X(x), \quad x \to \infty. \tag{8.141}$$

4. Then given the assumption that $\overline{F}_X \in RV_{-\alpha}$ for $\alpha > 1$ and the subsequent fact that $\overline{F}_Z \sim C\overline{F}_X$, one has that $\overline{F}_Z \in RV_{-\alpha}$ for $\alpha > 1$. In addition, this means that for any tail distribution such as $\overline{F}_X \in RV_{-\alpha}$, the random variable with this distribution has finite moments as follows

$$\mathbb{E}[X^\beta] < \infty, \quad \forall \beta < \alpha, \tag{8.142}$$

because of the property of RV distributions. This will also, therefore, apply for the annual loss.

5. Now utilizing the result in Lemma 8.2, where the asymptotic equivalence between two distribution tails up to a constant of proportionality is shown to produce under measurable functions the asymptotic equivalence of the integrals leading to the result in this context that because $\overline{F}_Z \sim \mathbb{E}[N]\overline{F}_X$, one has constant $C = \mathbb{E}$ and for the ES, one considers a function $\psi(x) = x$, the following asymptotic equivalence of the tail expectation applies

$$\int_{q_\kappa}^{\infty} x dF_Z(x) \sim \mathbb{E}[N] \int_{q_\kappa}^{\infty} x dF_X(x), \quad \text{as} \quad q_\kappa \to \infty, \tag{8.143}$$

where typically one sets $q_\kappa := \mathrm{VaR}_\kappa[Z]$.

6. Recall that one has a first-order approximation for the VaR of the annual loss which gives the following asymptotic equivalence for sub-exponential severity models

$$q_\kappa := \mathrm{VaR}_\kappa[Z] \sim F_X^{\leftarrow}\left(1 - \frac{1-\kappa}{E[N]}\right), \quad \kappa \uparrow 1. \tag{8.144}$$

7. Hence, the ES which was given by

$$\mathrm{ES}_\kappa[Z] = \frac{1}{1-\kappa}\int_{q_\kappa}^{\infty} x dF_Z(x), \tag{8.145}$$

with $q_\kappa := \mathrm{VaR}_\kappa[Z]$ can be rewritten as asymptotically equivalent to the following integral with respect to the severity model distribution instead of the annual loss distribution

$$\mathrm{ES}_\kappa[Z] \sim \frac{E[N]}{1-\kappa}\int_{q_\kappa}^{\infty} x dF_X(x). \tag{8.146}$$

8. Now using the fact that a severity density was assumed to be available, one can write

$$\mathrm{ES}_\kappa[Z] \sim \frac{E[N]}{1-\kappa}\int_{q_\kappa}^{\infty} x f_X(x) dx. \tag{8.147}$$

Furthermore, the fact that the distribution tail is regularly varying and the monotone density theorem can then be utilized to obtain the representation for the density $f_X(x) \sim \alpha x^{-\alpha-1}L(x)$ which produces using Proposition 8.17 the asymptotic equivalence representation

$$\mathrm{ES}_\kappa[Z] \sim \frac{E[N]}{1-\kappa}\alpha\int_{q_\kappa}^{\infty} x^{-\alpha}L(x) dx, \quad \kappa \uparrow 1. \tag{8.148}$$

9. Now given the assumption that $\alpha > 1$, one can utilize Karamata's theorem (see Chapter 3) which states that

$$\int_x^{\infty} s^\rho L(s) ds \sim \frac{-1}{\rho+2} x^{\rho+1} L(x), \quad x \to \infty. \tag{8.149}$$

Here one uses $\rho = -\alpha$ to obtain the following asymptotic equivalence for the ES as $\kappa \uparrow 1$ given by

$$\begin{aligned}\mathrm{ES}_\kappa[Z] &\sim \frac{E[N]}{1-\kappa}\alpha\int_{q_\kappa}^{\infty} x^{-\alpha}L(x) dx \\ &\sim \frac{E[N]}{1-\kappa}\frac{\alpha}{\alpha-1} q_\kappa^{1-\alpha} L(q_\kappa), \quad \kappa \uparrow 1.\end{aligned} \tag{8.150}$$

10. Now using the fact that $\overline{F} \in RV_{-\alpha}$ and using the quantile level q_κ given by the asymptotic equivalence

$$q_\kappa := \mathrm{VaR}_\kappa[Z] \sim F_X^{\leftarrow}\left(1 - \frac{1-\kappa}{E[N]}\right), \quad \kappa \uparrow 1, \tag{8.151}$$

one may consider the composition result in Proposition 8.18 to obtain by setting $\rho = -\alpha$ the asymptotic equivalence for the tail distribution composite given by

$$\overline{F}_X(q_\kappa) \sim \overline{F}_X\left(F^{\leftarrow}\left(1 - \frac{1-\kappa}{E[N]}\right)\right), \quad \kappa \uparrow 1. \tag{8.152}$$

11. One then has the following ES asymptotic equivalence that will form the first-order approximation for the LDA model given as $\kappa \uparrow 1$ by replacing the monomial term multiplying the slowly varying function by a known function given by the tail function for the severity $\overline{F}_X$ to obtain

$$\mathrm{ES}_\kappa[Z] \sim \frac{E[N]}{1-\kappa}\frac{\alpha}{\alpha-1} q_\kappa \overline{F}_X(q_\kappa). \tag{8.153}$$

This can then be further refined using the representation for q_κ as the VaR from the first-order approximation and the result in Equation 8.152 that produces the asymptotic equivalence

$$\mathrm{ES}_\kappa[Z] \sim \frac{E[N]}{1-\kappa}\frac{\alpha}{\alpha-1}F_X^{\leftarrow}\left(1-\frac{1-\kappa}{E[N]}\right)\overline{F}_X\left(F_X^{\leftarrow}\left(1-\frac{1-\kappa}{E[N]}\right)\right). \tag{8.154}$$

Simplifying this result using the continuity of the severity distribution, one obtains

$$\mathrm{ES}_\kappa[Z] \sim \frac{\alpha}{\alpha-1}F_X^{\leftarrow}\left(1-\frac{1-\kappa}{E[N]}\right). \tag{8.155}$$

Finally, this result is refined by showing it is asymptotically equivalent to a scaling of the first-order VaR expression to obtain the first-order SLA for the ES.

The ES can then be approximated for OpRisk settings up to first order of approximation via the result in Proposition 8.19.

Proposition 8.19 (First-Order SLA for Expected Shortfall) *Consider an LDA model at a fixed time $t > 0$. We assume that the distribution tail $\overline{F}$ of the severities is regularly varying with index $-\alpha$ for $\alpha > 1$ with an ultimately decreasing Lebesgue density f that is decreasing on (z, ∞) for a $z > 0$. Moreover, we assume that there exists an $\varepsilon > 0$ such that*

$$\sum_{n=0}^{\infty}(1+\varepsilon)^n \mathbb{P}\mathrm{r}[N=n] < \infty. \tag{8.156}$$

Then under these assumptions, one can derive the following asymptotic equivalence between the ES and the tail of the severity distribution given by

$$\mathrm{ES}_\kappa[Z] \sim \frac{\alpha}{\alpha-1}F_X^{\leftarrow}\left(1-\frac{1-\kappa}{\mathbb{E}[N]}\right) \sim \frac{\alpha}{\alpha-1}\mathrm{VaR}_\kappa[Z], \tag{8.157}$$

as $\kappa \uparrow 1$.

In the following two examples, we demonstrate the accuracy of the first-order approximation for the SLA for the risk measure ES under both the Poisson-Pareto and a Poisson-IG LDA families of models for a range of parameters, see derivations in Biagini and Ulmer (2009).

EXAMPLE 8.8 First-Order Approximation SLA for ES: Pareto Severity

Consider an LDA model with frequency given by a Poisson distribution $N \sim Poisson(\lambda)$ and a severity distribution for i.i.d. losses $X_i \sim F_X(x)$ which has tail function $\overline{F}_X$ which is regularly varying with parameter α such that $\overline{F}_X \in RV_{-\alpha}$ and a Lebesgue measurable density which is ultimately decreasing as given by the family of severity models obtained under the Pareto distribution function

$$F_X(x) = 1 - \left(1+\frac{x}{\theta}\right)^{-\alpha}, \quad x > 0, \tag{8.158}$$

for parameters $\alpha > 0$ and $\theta > 0$. The first-order asymptotic SLA for the ES is given by

$$\mathrm{ES}_\kappa[Z] \sim \frac{\alpha}{\alpha-1}\theta\left(\frac{E[N]}{1-\kappa}\right)^{\frac{1}{\alpha}}, \quad \kappa \uparrow 1. \tag{8.159}$$

■

EXAMPLE 8.9 Single Risk LDA Poisson-Inverse-Gaussian Family

Consider the heavy-tailed severity model, selected to model the sequence of i.i.d. losses in each year t, denoted by $\{X_i(t)\}_{i=1:N_t}$, and chosen to be an IG distribution $X_i \sim InverseGaussian(\widetilde{\mu}, \widetilde{\lambda})$ where the two parameters in this model correspond to parameterizing the mean of the distribution for the severity $\widetilde{\mu}$ and the shape of the distribution $\widetilde{\lambda}$. The survival and quantile functions of the severity are given by

$$f(x;\widetilde{\mu},\widetilde{\lambda}) = \left(\frac{\widetilde{\lambda}}{2\pi}\right)^{1/2} x^{-3/2} \exp\left\{-\frac{\widetilde{\lambda}(x-\widetilde{\mu})^2}{2\widetilde{\mu}^2 x}\right\},$$

$$F(x;\widetilde{\mu},\widetilde{\lambda}) = \Phi\left(\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}}-1\right)\right) + \exp\left(\frac{2\widetilde{\lambda}}{\widetilde{\mu}}\right)\Phi\left(-\sqrt{\frac{\widetilde{\lambda}}{x}}\left(\frac{x}{\widetilde{\mu}}+1\right)\right).$$

Therefore, the distribution for the annual loss $Z = \sum_{n=1}^{N} X_i$ is given in the closed-form LDA models in Chapter 6 where the IG family is closed under convolution, that is,

$$X_i \sim InverseGaussian(\widetilde{\mu},\widetilde{\lambda}) \Rightarrow S_n = \sum_{i=1}^{n} X_i \sim InverseGaussian(n\widetilde{\mu}, n^2\widetilde{\lambda}).$$

From this one can calculate the distribution of the compound process analytically and therefore approximate the ES risk measure for the partial sum via the approximation result developed in Example 8.1 Equation 8.24 with the new parameters for shape and scale. The drawback of this family is that there is no closed form for the inverse CDF, which obliges us to resort to a numerical procedure for obtaining the quantiles; fortunately, this is efficient and accurate for this class of models. For this model, we also present first- and second-order approximations for the VaR and SRM for different choices of parameters in Figures 8.5 and 8.6. ■

It is clear from the discussion above that the continuity and regular variation of the severity model was critical when combined with Karamata's theorem to obtain the first-order asymptotic for the ES SLA.

8.8.2 SECOND-ORDER SLA FOR EXPECTED SHORTFALL FOR REGULARLY VARYING LOSS MODELS

In deriving the second-order approximation, it will be useful to consider severity models which also satisfy second-order regular variation as discussed in Chapter 3. The reason this additional assumption is considered is that one may extend Karamata's theorem to a second-order regular

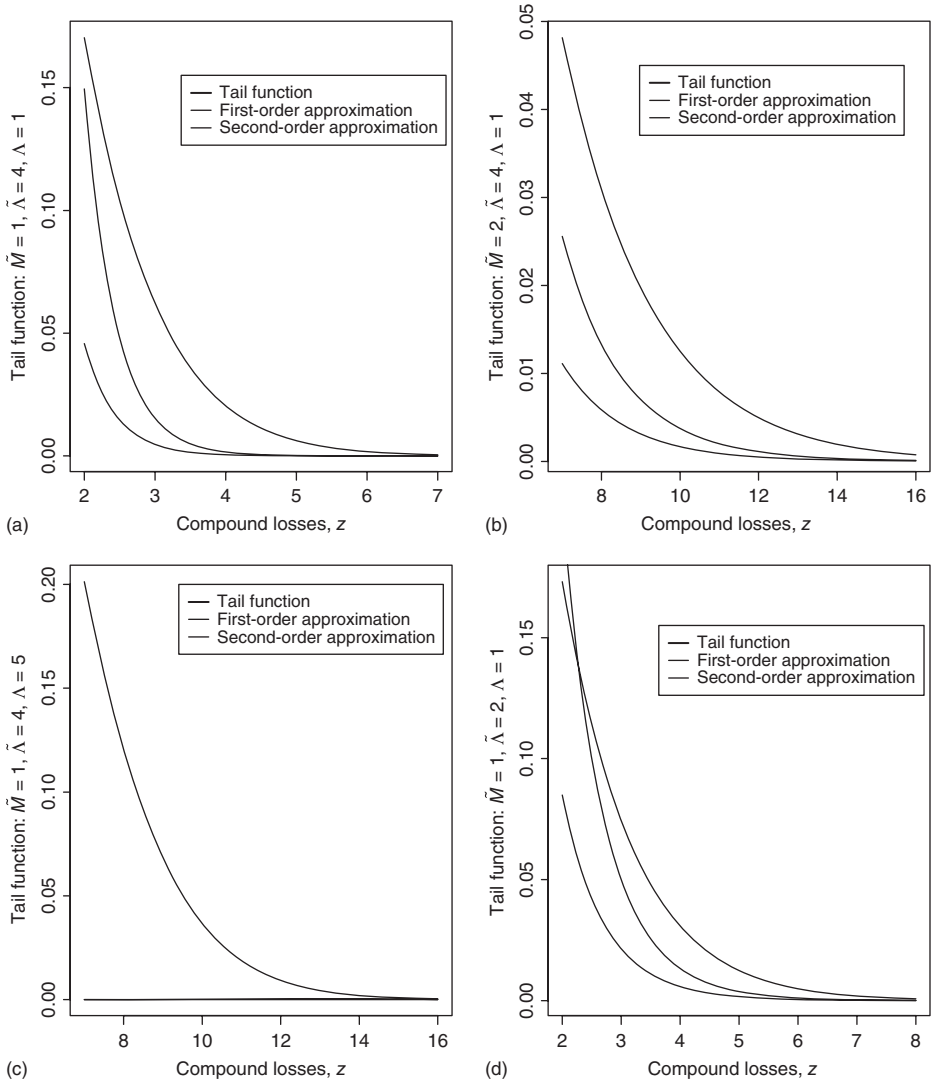

FIGURE 8.7 **Tail function approximation for the Poisson-inverse-Gaussian example.**

variation condition when the index of first-order regular variation $-\alpha < -1$. In Hua and Joe (2011), it was argued that such a result was particularly relevant for models that admit closed-form density but not necessarily closed-form distribution functions, see Proposition 8.20. It will first be relevant to recall the definition of second-order regular variation discussed in significantly more detail in Chapter 3, and for instance, Geluk *et al.* (1997).

Definition 8.12 (Second-Order Regular Variation) *The tail of a distribution function F with support on $[0, \infty)$ has a tail function $\overline{F}$ which is second-order regularly varying with first index $-\alpha$*

for $\alpha > 0$ and second index ρ denoted by $\overline{F} \in 2RV_{-\alpha,\rho}$ if there exists a function $A(x) \to 0$ as $x \to \infty$ which ultimately has constant sign such that the following holds:

$$\lim_{x\to\infty} \frac{1}{A(x)} \left[\frac{\overline{F}(x\lambda)}{\overline{F}(x)} - \lambda^{-\alpha} \right] = H(\lambda) := c\lambda^{-\alpha} \int_1^{\lambda} u^{\rho-1} du, \quad \lambda > 0, \tag{8.160}$$

for $c > 0$.

Remark 8.18 *In the case that $\lambda > 0$, one has the following forms for function $H(\lambda)$ given by*

$$H(\lambda) = \begin{cases} c\lambda^{-\alpha} \ln\lambda, & \text{if } \rho = 0, \\ c\lambda^{-\alpha} \dfrac{\lambda^{\rho} - 1}{\rho}, & \text{if } \rho < 0. \end{cases} \tag{8.161}$$

This then allows one to make the following useful extension to the representation offered by Karamata's theorem.

Proposition 8.20 (Karamata's Theorem for Second-Order Regular Variation) *Consider a function $f \in 2RV_{-\alpha,\rho}$ for $\alpha > 1$ and $\rho < 0$ with an auxiliary function $A(t)$. Then define $f^*(t) = \int_t^{\infty} f(x)dx$ and note that $f^* \in 2RV_{-\alpha+1,\rho}$ with an auxiliary function $A^*(t) = \frac{\alpha-1}{\alpha-1-\rho} A(t)$. Given these, one can show the following asymptotic equivalence holds between the integrated function and the original function as $t \to \infty$ given by*

$$f^*(t) \sim \frac{t}{\alpha - 1} f(t) + \frac{tA(t)}{\rho} \left(\frac{1}{\alpha - \rho - 1} - \frac{1}{\alpha - 1} \right) f(t). \tag{8.162}$$

This result will now be utilized to develop the second-order asymptotics for the ES risk measure based on the assumptions utilized in Hua and Joe (2011) that the severity distribution is not only regularly varying with index $-\alpha$ but also second-order regularly varying, see discussion in Chapter 3. Such assumptions allow one to utilize the following representation of the tail distribution in Proposition 8.21, see Hua and Joe (2011, lemma 3).

Proposition 8.21 (Representations of Second-Order Regularly Varying Distribution Tails) *Consider the tail of a distribution $\overline{F}$ that is second-order regularly varying $\overline{F} \in 2RV_{-\alpha,\rho}$, for $\alpha > 0$ and $\rho < 0$. Then one may express the tail of the distribution in the following form $\overline{F}(x) = kx^{-\alpha} L(x)$, for $x > 0$ with a constant $k > 0$ and the limit $\lim_{x\to\infty} L(x) = 1$ with $|1 - L(x)| \in RV_{\rho}$.*

Using this representation, one can derive the following second-order asymptotic expansion for the ES in Proposition 8.22, which is in terms of the annual loss VaR, for which we can again replace the true value with an asymptotic expansion from above if required, see details in Hua and Joe (2011, proposition 5).

Proposition 8.22 (Second-Order SLA for the ES Under Second-Order RV) *Consider a non-negative random variable Z which has a distribution which satisfies $\overline{F}_Z \in 2RV_{-\alpha,\rho}$ for $\alpha > 1$ and $\rho < 0$. Then the tail distribution $\overline{F}_Z(x)$ can be written as $\overline{F}_Z(x) = kx^{-\alpha} L(x)$ and the following asymptotic equivalence applies for the ES risk measure at a VaR level $\mathrm{VaR}_{\kappa}[Z]$ as $x \to \infty$, that is, the quantile level $\kappa \uparrow 1$ giving*

$$\mathrm{ES}_{\kappa}[Z] \sim \frac{\alpha}{\alpha - 1}\mathrm{VaR}_{\kappa}[Z] + \nu(\mathrm{VaR}_{\kappa}[Z]), \tag{8.163}$$

with the function $\nu(\cdot)$ *given by*

$$\nu(x) = \frac{xa(x)}{(\alpha - 1 - \rho)(\alpha - 1)}, \tag{8.164}$$

and $a(x) = -\rho\frac{1-L(x)}{L(x)}$.

To complement this second-order asymptotic result, one has the bound obtained in de Haan and Ferreira (2006, theorem 2.3.9) which provides a uniform inequality for second-order regularly varying tail distribution functions, see Proposition 8.23.

Proposition 8.23 (Uniform Bounds on Second-Order RV Distribution Tails) *Consider the tail distribution* $\overline{F}_Z$ *that is assumed to be second-order regularly varying* $\overline{F}_Z \in 2RV_{-\alpha,\rho}$ *with* $\alpha > 0$ *and* $\rho < 0$. *Then for any* $\varepsilon > 0$ *and* $\delta > 0$, *there exists a constant* $x_0 = x_0(\varepsilon, \delta)$ *s.t. For all* x, *one has* $x\lambda \geq x_0$ *and* $\lambda > 0$ *which satisfies the following inequality uniformly:*

$$\left| \frac{1}{a(x)} \left[\frac{\overline{F}_Z(\lambda x)}{\overline{F}_Z(x)} - \lambda^{-\alpha} \right] - \lambda^{-\alpha} \frac{\lambda^{\rho} - 1}{\rho} \right| \leq \varepsilon \max\{\lambda^{-\alpha+\rho+\delta}, \lambda^{-\alpha+\rho-\delta}\}, \tag{8.165}$$

where $a(t) = -\rho\frac{[1-L(x)]}{L(x)}$ *with* $\overline{F}_Z(x) = kx^{-\alpha}L(x)$, *for* $k \in (0, \infty)$ *and* $\lim_{x\to\infty} L(x) = 1$.

As noted in Hua and Joe (2011), the second-order expansion of the ES risk measure in terms of the function

$$\nu(x) = \frac{xa(x)}{(\alpha - 1 - \rho)(\alpha - 1)} \tag{8.166}$$

shows that the second-order term becomes increasingly prominent as $\rho > -1$ because $|xa(x)|$ increases as $x \to \infty$. They, therefore, note that the second-order term becomes relevant when $\rho \in (-1, 0)$.

It should also be noted that the tail of the annual loss distribution cannot be too heavy tailed, as otherwise the ES will be infinite. In terms of the parameters of the RV, one can say that the ES is well defined when $\alpha > 1$, because for $\alpha < 1$, one has that the conditional tail expectation is infinite for any quantile level $\kappa < 1$.

8.8.3 EMPIRICAL PROCESS AND EVT APPROXIMATIONS OF EXPECTED SHORTFALL

We will start this section discussing classical EVT-based estimators for the ES or TCE and then move to empirical-process-based estimators that draw on aspects of EVT to develop estimators for the TCE.

EVT-based SLA for the ES have also been developed such as the method of Marinelli *et al.* (2007) which is based on a POTs framework, see background discussion in Chapter 2. The first result to note is that if the severity distribution is assumed to be from the GEV family such that each loss is i.i.d. with $X_i \sim G_{\gamma,\sigma}$, then for $\gamma < 1$ and $\sigma + \gamma x > 0$, one has the following estimator for the TCE given by

$$\mathbb{E}[X|X > q_\kappa] = \frac{q_\kappa + \sigma}{1-\gamma}. \tag{8.167}$$

If one assumes that the exceedance distribution for each $X_i - u$ given $X_i > u$ follows a GPD model with parameters γ and $\sigma + \gamma(x_p - u)$, then one has the TCE given by

$$\mathrm{ES}_\kappa[Z] = \mathbb{E}[X|X > \mathrm{VaR}_\kappa[Z]] = \frac{\mathrm{VaR}_\kappa[Z]}{1-\gamma} + \frac{\sigma - \gamma u}{1-\gamma} \tag{8.168}$$

which can be estimated by simply estimation or approximation of the VaR as discussed earlier and estimation of the GPD model parameters as discussed in detail in Chapter 2.

Next we discuss empirical-process-based estimators for the TCE which utilize EVT. In discussing the point-process-based estimators for ES, it should first be noted that under the assumption that the second moment of the annual loss distribution is finite, then the asymptotic normality of the non-parametric conditional tail estimator TCE has been shown to hold, see details in Brazauskas *et al.* (2008). Importantly, in OpRisk settings where heavy-tailed loss processes are of significance, this result will no longer be applicable. In Necir *et al.* (2010), they developed estimators based on point process results from EVT for models which admit a finite mean and infinite variance for the annual loss model.

If one assumes that the annual loss distribution LDA model F_Z is unknown but assumed to be continuous, then one may adopt a point-process-based estimation framework for approximation of the TCE. First, the results for the light-tailed finite mean and finite variance estimators will be considered before moving to the heavier-tailed case where the variance is no longer finite.

In the lighter-tail context, one will assume that there are individual losses $\{X_i\}_{i=1}^n$ which are assumed i.i.d. with distribution F_X that admits finite mean and variance. The order statistics from these losses will be denoted by the set of random variables from smallest to largest given by $\{X_{(i,n)}\}_{i=1}^n$. Then the simplest estimator of the ES is given by using the empirical quantile function to estimate

$$\widehat{\mathrm{TCE}}_{\kappa,n}[Z] = \frac{1}{1-\kappa}\int_\kappa^1 \widehat{Q}_n(s)\,ds, \tag{8.169}$$

where $\widehat{Q}_n$ is the empirical estimate of the quantile function obtained from the order statistics directly. In Brazauskas *et al.* (2008), they studied the asymptotic properties of this simple TCE estimator and the following convergence in distribution result was obtained, as shown in Propositions 8.24 and 8.25.

Proposition 8.24 *Consider a loss process with finite first moment for the severity distribution* $\mathbb{E}[X] < \infty$. *Then for every* $\kappa \in [0,1]$, *one can show that the corresponding empirical estimate of the conditional tail expectation* $\widehat{\mathrm{TCE}}_{\kappa,n}[Z]$ *converges to the true conditional tail expectation* $\mathrm{TCE}_\kappa[Z]$] *almost surely and, therefore, in probability. Hence,* $\widehat{\mathrm{TCE}}_{\kappa,n}[Z]$ *is both a strongly and a weakly consistent estimator of* $\mathrm{TCE}_\kappa[Z]$ *for every* $\kappa \in [0,1]$.

Remark 8.19 *It is, however, easy to show that this empirical TCE estimator is not a uniformly consistent estimator of the true TCE over the entire interval* $[0,1]$, *see details in Brazauskas* et al. *(2008).*

Proposition 8.25 (Convergence in Distribution of Empirical TCE Estimator) *If the second moment of the loss distribution is finite* $\mathbb{E}[X^2] < \infty$*, then for every* $\kappa \in (0,1)$*, one has the following convergence in distribution applying as the sample size goes to infinity* ($n \to \infty$)*,*

$$\sqrt{n}(\widehat{\mathrm{TCE}}_{\kappa,n}[Z] - \mathrm{TCE}_{\kappa}[Z])(1-\kappa) \overset{d}{\Rightarrow} Normal(0, \sigma^2(\kappa)), \tag{8.170}$$

where $\mathrm{TCE}_{\kappa}[Z]$ *represents the unknown true conditional tail expectation for level* κ *and the asymptotic variance given by*

$$\sigma^2(\kappa) = \int_{\kappa}^{1}\int_{\kappa}^{1} (\min(x,y) - xy)\, dQ(x)\, dQ(y). \tag{8.171}$$

Using this result, one can of course construct a confidence interval for the empirically estimated TCE as follows:

$$\left[\widehat{\mathrm{TCE}}_{\kappa,n}[Z] - \frac{1}{\sqrt{n}} z_{\frac{\alpha}{2}} \widehat{\sigma}_n(\kappa), \widehat{\mathrm{TCE}}_{\kappa,n}[Z] + \frac{1}{\sqrt{n}} z_{\frac{\alpha}{2}} \widehat{\sigma}_n(\kappa)\right] \tag{8.172}$$

with $z_{\frac{\alpha}{2}}$ the standard Gaussian z-score and the variance estimated by

$$\widehat{\sigma}_n^2(\kappa) = \frac{1}{(1-\kappa)^2} \sum_{n\kappa \le j \le (n-1)} \sum_{n\kappa \le k \le (n-1)} \left(\frac{j \wedge k}{n} - \frac{j}{n}\frac{k}{n}\right) \left(X_{(j+1,n)} - X_{(j,n)}\right)\left(X_{(k+1,n)} - X_{(k,n)}\right), \tag{8.173}$$

see details in Brazauskas *et al.* (2008).

As noted in Necir *et al.* (2010), the assumptions in the above results can be limiting in their application, especially in OpRisk settings. For instance, in the case of the Pareto severity model with $\gamma \in (0,1)$ so that the TCE exists, because for $\gamma > 1$, the conditional moments are infinite because of the extremely heavy tails. If the above assumptions are imposed on the finite second moment, then this would only be applicable if the tail index parameter $\gamma \in \left(0, \frac{1}{2}\right)$ which is restrictive in practice. To relax this assumption to allow for $\gamma \in (0,1)$ and, therefore, to obtain the TCE estimator for such cases a modification to the above estimator needs to be developed.

Such an estimator was developed by Necir *et al.* (2010) and is given in Equation 8.174, and it explicitly utilizes the Hill estimator of the EVI parameter for the severities to modify the TCE estimate so that it is well defined when $\mathbb{E}[X^2]$ is not finite. This estimator is based on the different methods one ought to adopt when estimating high quantiles with heavy-tailed models, see discussions in Weissman (1978), Dekkers and De Haan (1989) and Matthys and Beirlant (2003).

$$\widetilde{\mathrm{TCE}}_{\kappa,n}[Z] = \frac{1}{1-\kappa}\int_{\kappa}^{1-\frac{k}{n}} \widehat{Q}_n(s)\, ds + \frac{kX_{(n-k,n)}}{n(1-\kappa)(1-\widehat{\gamma})} \tag{8.174}$$

with the Hill estimator used for the EVI parameter estimate given by

$$\widehat{\gamma} = \frac{1}{k}\sum_{i=1}^{k} \ln X_{(n-i+1,n)} - \ln X_{(n-k,n)}. \tag{8.175}$$

One should select the integers $k \in \{1, 2, \ldots, n\}$ such that as $n \to \infty$ one has $k \to \infty$ and $\frac{k}{n} \to 0$.

This new estimator is obtained by splitting the integral from $[\kappa, 1]$ into two parts, one for $\left[\kappa, 1 - \frac{k}{n}\right]$ which resembles that standard TCE and which is estimated based on the standard quantile function and then integral from $\left[1 - \frac{k}{n}, 1\right]$ which is based on the high quantile estimator given by Weissman for heavy-tailed models

$$\widehat{Q}_n(s) = \left(\frac{k}{n}\right)^{\widehat{\gamma}} X_{(n-k,n)} s^{-\widehat{\gamma}}, \quad s \in \left[0, \frac{k}{n}\right] \tag{8.176}$$

with $\widehat{\gamma}$ the estimated EVI parameter via say the Hill estimator as above or the Pickands estimator, see discussion in Chapter 2. It is the integration of this second quantile estimator over the domain $\left[1 - \frac{k}{n}, 1\right]$ that yields the second term in the TCE estimator.

In Necir *et al.* (2010), the following convergence in distribution is shown, given Proposition 8.26 applies to the new estimator.

Proposition 8.26 (Convergence in Distribution of Empirical Modified TCE Estimator) *Consider a severity distribution F_X which satisfies the condition for second-order regular variation with second parameter $\rho \leq 0$ and there exists a function $a(x)$ that does not change sign in the neighbourhood of infinity such that for all $\lambda > 0$, the severity distribution satisfies*

$$\lim_{x \to \infty} \frac{1}{a(x)} \left(\frac{1 - F_X(x\lambda)}{1 - F_X(x)} - \lambda^{-\frac{1}{\gamma}} \right) = \lambda^{-\frac{1}{\gamma}} \frac{\lambda^{\frac{\rho}{\gamma}} - 1}{\frac{\rho}{\gamma}} \tag{8.177}$$

for $\gamma \in \left(\frac{1}{2}, 1\right)$. Then, for any sequence of integers $k = k_n \to \infty$ such that $\frac{k}{n} \to 0$ and $\sqrt{k} A\left(\frac{k}{n}\right) \to 0$ as $n \to \infty$, one has for any fixed $\kappa \in (0, 1)$ the convergence in distribution for the modified empirical TCE estimator $\widetilde{\mathrm{TCE}}_{\kappa,n}[Z]$ given by

$$\frac{\sqrt{n}(\widetilde{\mathrm{TCE}}_{\kappa,n}[Z] - \mathrm{TCE}_{\kappa}[Z])}{\left(\frac{k}{n}\right)^{\frac{1}{2}}} X_{(n-k,n)} \stackrel{d}{\Rightarrow} Normal(0, \sigma_\gamma^2), \tag{8.178}$$

where the asymptotic variance σ_γ^2 is given by

$$\sigma_\gamma^2 = \frac{\gamma^4}{(1-\gamma)^4(2\gamma - 1)}. \tag{8.179}$$

Remark 8.20 *It is interesting to observe that the resulting asymptotic variance is no longer a function of κ, see discussion Necir* et al. *(2010).*

Again one can, therefore, utilize this result to obtain confidence intervals for the modified empirical estimator for the TCE given by

$$\left[\widetilde{\mathrm{TCE}}_{\kappa,n}[Z] - z_{\frac{\alpha}{2}} \frac{\left(\frac{k^*}{n}\right)^{\frac{1}{2}} X_{(n-k^*,n)} \sigma^2_{\widehat{\gamma}_n}}{(1-\kappa)\sqrt{n}}, \widetilde{\mathrm{TCE}}_{\kappa,n}[Z] + z_{\frac{\alpha}{2}} \frac{\left(\frac{k^*}{n}\right)^{\frac{1}{2}} X_{(n-k^*,n)} \sigma^2_{\widehat{\gamma}_n}}{(1-\kappa)\sqrt{n}} \right] \tag{8.180}$$

for k^* selected according to the method proposed in Cheng and Peng (2001).

8.8.4 SLA FOR SPECTRAL RISK MEASURES

Since the work of Acerbi (2002) where the space of coherent risk measures was obtained via expansions of coherent elementary basis measures, there has been a significant consideration in the actuarial and risk literature regarding families of SRMs. Working with a family of SRMs requires the OpRisk quantitative risk manager to specify or elicit from business experts a measure of risk aversion. In Acerbi (2002), the necessary and sufficient conditions for the risk-aversion function, denoted here generically by ϕ, was developed to ensure the resulting family of SRM that is formed from this risk-aversion function produces coherent risk measures. Basically, the studied question involved exploring under what conditions on the weighting function can one consider a weighted tail expectation to produce a family of coherent risk measures. To address this inquiry, they utilized the fact discussed in Cruz *et al.* (2015, chapter 6) that given known risk measures that are convex, one may construct new risk measures which preserve the coherency through a convex linear combination. That is, given k risk measures, denoted generically here by $\rho_1, \rho_2, \ldots, \rho_k$, one can construct a new risk measure $\widetilde{\rho} = \sum_{i=1}^{k} w_i \rho_i$ where the weights satisfy $w_i \in \mathbb{R}^+$ and normalization condition $\sum_{i=1}^{k} w_i = 1$. Given this, one may then formulate an understanding of the properties of the space of coherent measures that are generated by the most general convex combination of κ-ES.

To proceed with an understanding of the class of SRM and risk-aversion functions, we first make a few basic definitions that will aid in the understanding of the relationship between the SRM and the risk-aversion function ϕ as described in detail in Acerbi (2002). Starting with the definition of a positive and decreasing risk-aversion function element which satisfies the following conditions making it a valid admissible risk spectrum given in Definition 8.13, see Acerbi (2002).

Definition 8.13 [Conditions for an Admissible Risk Spectrum] *An element function on the normed space $\mathscr{L}^1$ given by $\phi \in \mathscr{L}^1([a,b])$ is termed an admissible risk spectrum if it satisfies the following three conditions:*

1. *The element ϕ satisfies for all interval subsets $I \subset [a,b]$ the following integral positivity*

$$\int_I \phi(s)\,ds \geq 0, \tag{8.181}$$

   that is, it is a positive element.

2. *The element ϕ satisfies for values $q \in (a,b)$ and for all $\varepsilon > 0$ satisfying $[q-\varepsilon, q+\varepsilon] \subset [a.b]$ the following integral*

$$\int_{q-\varepsilon}^{q} \phi(s)\,ds \geq \int_{q}^{q+\varepsilon} \phi(s)\,ds, \tag{8.182}$$

   that is, it is a decreasing element.

3. *The element ϕ is normalized such that*

$$\|\phi\| = \int_a^b |\phi(s)|\,ds = 1. \tag{8.183}$$

Using the definition of an admissible risk spectrum function, one can define the corresponding risk measure generated by the risk spectrum function as given in Proposition 8.27, see Acerbi (2002, theorem 2.5)

Proposition 8.27 (Risk Measure Generated by Admissible Risk Spectrums) *Consider the admissible risk spectrum function $\phi \in \mathscr{L}^1([0,1])$ also known as the risk-aversion function that satisfies the three conditions in Definition 8.13. Then the following integral defines the SRM denoted by $M_\phi(X)$ for loss random variable $X \sim F_X$ that is generated by this risk spectrum according to the relationship,*

$$M_\phi[X] = -\int_0^1 F_X^{\leftarrow}(s)\phi(s)\,ds. \tag{8.184}$$

It is important to note that under the above framework, it is critical that the risk-aversion function satisfies admissibility for the resulting SRM generated to be a valid risk measure.

At this stage, it is also relevant to note that a different approach was adopted in Kusuoka (2001) to study the class of spectral measures where they proved that the spectral measures can also be characterized as the coherent measures which are law-invariant and comonotonic additive. Furthermore, it was shown in Mainik and Rüschendorf (2010) that for any admissible risk spectrum $\phi : [0,1] \mapsto \mathbb{R}$, the transformation $\tau_\kappa : s \mapsto 1 - \frac{1-s}{1-\kappa}$ for $\kappa \in (0,1)$ and $s \in [\kappa, 1]$ produces a rescaled new class of admissible risk spectra $\phi_\kappa(s)$ given by the rescaling

$$\begin{aligned}\phi_\kappa(s) &= \tau_\kappa'(s)\phi(\tau_\kappa(s))\mathbb{I}_{[0,1]}(\tau_\kappa(s)) \\ &= \frac{1}{1-\kappa}\phi\left(1 - \frac{1-s}{1-\kappa}\right)\mathbb{I}_{[\kappa,1]}(s).\end{aligned} \tag{8.185}$$

Given this new rescaled risk-aversion function, one then obtains the OpRisk SRM according to

$$\mathrm{SRM}_{\phi_\kappa}[Z] = \int_\kappa^1 \phi_\kappa(s)\mathrm{VaR}_s[Z]\,ds. \tag{8.186}$$

Having defined the appropriate properties that a risk-aversion function must satisfy to produce a valid SRM, one can now turn to the asymptotic approximation of this risk measure for first- and second-order expansions, starting with the first-order result presented in Proposition 8.28, see Tong and Wu (2012, theorem 3.1).

Proposition 8.28 (First-Order SLA for Spectral Risk Measures) *Consider a severity distribution tail $\overline{F}_X$ that is regularly varying $\overline{F}_X \in RV_{-\alpha}$ for $\alpha > 1$ so that the ES is well defined. Then let ϕ_κ be an admissible risk-aversion function (spectrum) given*

$$\begin{aligned}\phi_\kappa(s) &= \tau_\kappa'(s)\phi(\tau_\kappa(s))\mathbb{I}_{[0,1]}(\tau_\kappa(s)) \\ &= \frac{1}{1-\kappa}\phi\left(1 - \frac{1-s}{1-\kappa}\right)\mathbb{I}_{[\kappa,1]}(s),\end{aligned} \tag{8.187}$$

with base admissible risk spectra ϕ satisfying the conditions for all $x \in (1,\infty)$ of the following upper bound

$$\phi\left(1 - \frac{1}{x}\right) \leq Kx^{-\frac{1}{\alpha}+1-\varepsilon}, \tag{8.188}$$

for some constants $K > 0$ and $\varepsilon > 0$. Furthermore, assume that the value of ε also satisfies the condition with respect to the frequency distribution that

$$\sum_{n=0}^{\infty}(1+\varepsilon)^n \mathbb{P}\mathrm{r}[N = n] < \infty. \tag{8.189}$$

Then one can state the following first-order asymptotic for the LDA models SRM given by, as $x \to \infty$,

$$\begin{aligned}\mathrm{SRM}_{\phi_\kappa}[Z] &\sim \mathscr{K}(\alpha,\phi)F_X^{\leftarrow}\left(1-\frac{1-\kappa}{\mathbb{E}[N]}\right)\\ &\sim \mathscr{K}(\alpha,\phi)\mathrm{VaR}_\kappa[Z],\end{aligned} \tag{8.190}$$

with $\mathscr{K}(\alpha,\phi)=\int_1^\infty s^{\frac{1}{\alpha}-2}\phi\left(1-\frac{1}{s}\right)ds$.

Remark 8.21 *To recover the first-order asymptotic for the ES discussed earlier from this SRM asymptotic, one can consider the function* $\phi(s)=1$ *for all* $s\in[0,1]$. *This will produce a new rescaled admissible risk spectrum defined by*

$$\phi_\kappa(s)=\begin{cases}0, & \text{if } s<\kappa,\\ \dfrac{1}{1-\kappa}, & \text{if } s\geq\kappa,\end{cases} \tag{8.191}$$

which will weight all tail quantiles for $s\geq\kappa$ *the same weight of* $\frac{1}{1-\kappa}$, *and in this case, the SRM associated to this admissible risk-aversion function* ϕ_κ *is the ES. This will also lead to the function* $\mathscr{K}(\alpha,\phi)=\frac{\alpha}{\alpha-1}$.

One of the more widely used risk-aversion functional forms involves the exponential utility function which produces a family of risk-aversion functions known as the CARA functions. In Example 8.10, the first-order SLA under the CARA function is detailed, see discussion in Tong and Wu (2012).

EXAMPLE 8.10 First-Order SLA for Spectral Risk Measures via CARA Functions

The CARA family of admissible risk aversion functions is given by

$$\phi(\kappa)=\frac{\zeta\exp(-\zeta(1-\kappa))}{1-\exp(-\zeta)}, \tag{8.192}$$

where ζ is given by the Arrow–Pratt coefficient of absolute risk aversion $\zeta=-\frac{U''(x)}{U'(x)}$ for exponential utility functions $U(x)=-\exp(-\zeta x)$. Now substituting this base CARA function into the modified risk-aversion function

$$\begin{aligned}\phi_\kappa(s)&=\tau_\kappa'(s)\phi(\tau_\kappa(s))\mathbb{I}_{[0,1]}(\tau_\kappa(s))\\ &=\frac{1}{1-\kappa}\phi\left(1-\frac{1-s}{1-\kappa}\right)\mathbb{I}_{[\kappa,1]}(s),\end{aligned} \tag{8.193}$$

one obtains the new admissible rescaled risk-aversion function which generates a spectral risk measure with the following first-order asymptotic relationship

$$\mathrm{SRM}_{\phi_\kappa}[Z]\sim\frac{\gamma^{\frac{1}{\alpha}}}{1-\exp(-\gamma)}\mathrm{VaR}_\kappa[Z]\int_0^\gamma u^{-\frac{1}{\alpha}}\exp(-u)\,du, \tag{8.194}$$

where $\mathscr{K}(\alpha,\phi)=\frac{\gamma^{\frac{1}{\alpha}}}{1-\exp(-\gamma)}\int_0^\gamma u^{-\frac{1}{\alpha}}\exp(-u)\,du$. ■

In Tong and Wu (2012), it was demonstrated that if one further imposes the following condition on the severity distribution in the LDA model such that the following second-order condition is satisfied for second-order regular variation,

$$\lim_{x\to\infty} \frac{1}{A(x)} \left[\frac{U_F(x\lambda)}{U_F(x)} - \lambda^{\frac{1}{\alpha}} \right] = \lambda^{\frac{1}{\alpha}} \frac{\lambda^{\rho} - 1}{\rho}, \tag{8.195}$$

where $A(t) \in \mathrm{RV}_{\rho}$ for $\rho \leq 0$, and $U_F(x) := \left(\frac{1}{1-F}\right)^{\leftarrow}(x) \in \mathrm{RV}_{1/\alpha}$. Basically, this is the equivalent condition used for the severity model in the second-order asymptotic derivation for the ES. This is not a surprise that the same result is utilized here for the SRM because the ES is a special case as has been discussed. The resulting second-order asymptotic result for the SRM is then given in Proposition 8.29.

Proposition 8.29 (Second-Order SLA for Spectral Risk Measures) *Consider a severity distribution tail $\overline{F}_X$ that is regularly varying $\overline{F}_X \in \mathrm{RV}_{-\alpha}$ for $\alpha > 1$ so that the ES is well defined. Then let ϕ_{κ} be an admissible risk-aversion function (spectrum) given*

$$\begin{aligned} \phi_{\kappa}(s) &= \tau_{\kappa}'(s)\phi(\tau_{\kappa}(s))\mathbb{I}_{[0,1]}(\tau_{\kappa}(s)) \\ &= \frac{1}{1-\kappa}\phi\left(1 - \frac{1-s}{1-\kappa}\right)\mathbb{I}_{[\kappa,1]}(s) \end{aligned} \tag{8.196}$$

with base admissible risk spectra ϕ satisfying the conditions for all $x \in (1, \infty)$ of the following upper bound

$$\phi\left(1 - \frac{1}{x}\right) \leq K x^{-\frac{1}{\alpha}+1-\varepsilon} \tag{8.197}$$

for some constants $K > 0$ and $\varepsilon > 0$. Furthermore, assume that the value of ε also satisfies the condition with respect to the frequency distribution that

$$\sum_{n=0}^{\infty} (1+\varepsilon)^n \mathbb{P}\mathrm{r}(N = n) < \infty. \tag{8.198}$$

Furthermore, one assumes that $U_F(x) := \left(\frac{1}{1-F}\right)^{\leftarrow}(x) \in \mathrm{RV}_{\frac{1}{\alpha}}$ is second-order regularly varying such that $U_F(x) \in 2\mathrm{RV}_{\frac{1}{\alpha},\rho}(A(x))$ with second-order parameter $\rho \leq 0$. Then one can state the following second-order asymptotic for the LDA models SRM given by, as $x \to \infty$,

$$\mathrm{SRM}_{\phi_{\kappa}}[Z] \sim \left[\mathscr{K}(\alpha, \phi) + A\left(\frac{\mathbb{E}[N]}{1-\kappa}\right) \mathscr{M}(\alpha, \phi, \rho) \right] F_X^{\leftarrow}\left(1 - \frac{1-\kappa}{\mathbb{E}[N]}\right) \tag{8.199}$$

with

$$\begin{aligned} \mathscr{K}(\alpha, \phi) &= \int_1^{\infty} s^{\frac{1}{\alpha}-2} \phi\left(1 - \frac{1}{s}\right) ds \\ \mathscr{M}(\alpha, \phi, \rho) &= \frac{1}{\rho} \int_1^{\infty} s^{\frac{1}{\alpha}-2}(s^{\rho} - 1)\phi\left(1 - \frac{1}{s}\right) ds. \end{aligned} \tag{8.200}$$

8.9 Assessing the Accuracy and Sensitivity of the Univariate SLA

As has been discussed earlier and mentioned in the discussed references, it is important to have a thorough understanding of the accuracy and applicability of the SLA approximation. The reason for this is that utilization of such risk measure approximations will result in financial institutions setting aside significant capital amounts to guard against adverse losses associated with a given risk type and business cell. In addition, the stability of the resulting financial institution and the stability of its estimated capital amounts over time will reflect the stability of the estimate of the VaR both at an institution wide level and at the level of individual risk type and business cells levels in a hierarchical business structure. Therefore, it is meaningful to carefully study the approximation error of the SLA and the practical sources of such error in the resulting capital estimations under this SLA.

In addition, we will demonstrate, via an assessment of the SLA approximation error, in what sense the SLA-VaR approximation refinements of Böcker and Sprittulla (2006) and Degen (2010) are actually improvements of the basic SLA-VaR approximation presented in Definition 8.9.

It is apparent that the estimation of the SLA-VaR will be affected by three main sources of error, two of which will be common to all VaR estimation approaches and one which will be specific to an SLA approach. We consider two of these sources of error with respect to the SLA approach and denote the decomposition of these three sources of error in the SLA of the annual loss VaR for a given risk type and business cell by the following generic notation,

$$e_\alpha(\text{SLA} - \text{VaR}_\alpha) = e_\alpha(\mathscr{M}) + e_\alpha(\widehat{\boldsymbol{\Theta}}) + e_\alpha(\text{SLA}). \tag{8.201}$$

We explicitly denote the dependence on each of these error terms as a function of the risk measure significance level α. This is important to note as some of these sources of error will decrease in magnitude asymptotically with $\alpha \to 1$, whilst other sources of error may increase in magnitude.

Hence, the total error $e_\alpha(\text{SLA} - \text{VaR}_\alpha)$ in an SLA-VaR estimate at a quantification level of α will depend on the following three sources:

1. *Model Structure Uncertainty.* $e_\alpha(\mathscr{M})$ which represents the model uncertainty associated with the choice of parametric or semi-parametric model utilized to represent both the severity process and the frequency process in the LDA structure. It, therefore, represents the model structure uncertainty, the likelihood of working with models with severe losses as possible outcomes and the relative severity of such losses. In addition, it may represent the chance of observing overdispersion effects in the frequency of occurrence of losses. In this regard, it is clear that such effects will dissipate as $\alpha \to 1$. However, in general, quantification of this model error is difficult to achieve with precision, although it is possible to achieve within a family of parametric models via standard statistical approaches which include hypothesis testing on the model choice, predictive performance criterion in and out of sample based on mean square errors of prediction (MSEP), information theoretic criterion such as corrected Akaike information criterion (AICc), Schwartz's criterion and in a Bayesian setting via Bayes factors, deviance information criterion (DIC) and Bayesian information criterion (BIC) criterion.

2. *Parameter Estimation Error.* $e_\alpha(\widehat{\boldsymbol{\Theta}})$ which represents the error associated with the parameter estimation, having made a choice of severity and frequency model to consider in the LDA framework. In many popular severity and frequency models adopted by OpRisk practitioners, the point estimator associated with the estimation of the model parameters, denoted by the random vector $\widehat{\boldsymbol{\Theta}}$, will be unbiased ($\mathbb{E}[\widehat{\boldsymbol{\Theta}}] = \boldsymbol{\theta}$), consistent as a function of the sample size n ($\mathrm{Var}(\widehat{\boldsymbol{\Theta}}) \to 0, \text{ as } n \to \infty$) and possibly efficient (minimum variance in the class of estimators considered). Such results will apply to maximum-likelihood estimators in many cases relevant to OpRisk LDA models. However, these properties, although statistically important, are often not practically relevant for OpRisk modeling. The reason for this is that the sample sizes associated with loss processes considered to have high impact but to occur infrequently or rarely in practice results in real-world modeling situations in which the associated asymptotics in the data sample size used in the parameter estimation are insignificant in size and so such asymptotic results are irrelevant in these settings. Even if such statistically relevant properties cannot be shown to hold, the estimation accuracy of the parameter estimates themselves can be quantified. This can be achieved in a classical frequentist setting via the specification of confidence interval (region) based on the sampling distribution of the parameter estimator $\widehat{\boldsymbol{\Theta}} \sim F_{\boldsymbol{\Theta}}(\theta)$ and the relevant data to the parameter estimation, denoted generically by $\mathbf{Y}$. The confidence interval with pivot $g(\boldsymbol{\Theta}, \mathbf{Y})$ is, therefore, given by

$$\mathbb{P}\mathrm{r}[L \leq g(\boldsymbol{\Theta}, \mathbf{Y}) \leq U] = 1 - \kappa,$$

which results after rearranging the pivot in a κ level confidence interval (region) realized for the estimated parameter vector given by $\widehat{\boldsymbol{\theta}} \in [l(\widehat{\boldsymbol{\theta}}, \mathbf{y}, \kappa), u(\widehat{\boldsymbol{\theta}}, \mathbf{y}, \kappa)]$. Alternatively, if one is working under a Bayesian estimation paradigm, parameter uncertainty can be quantified by the Bayesian equivalent of the confidence interval, the posterior credible interval, see discussion in Cruz *et al.* (2015, chapter 7). The quantification of these measures of parameter uncertainty can be easily obtained analytically or numerically via, for example, a boot-strap-based procedure, see Efron and Tibshirani (1993). What is not so trivial is the incorporation of this uncertainty associated with the point estimators of the parameters of the LDA model into the analysis of the error term $e_\alpha(\widehat{\boldsymbol{\Theta}})$. This will be discussed below.

3. *Single-Loss Approximation Error.* $e_\alpha(SLA)$ represents the error attributed to making the asymptotic tail approximation associated with the SLA-VaR defined in Theorem 7.7. As demonstrated in Example 8.3, this error can be significant even when the error terms $e_\alpha(\mathscr{M})$ and $e_\alpha(\widehat{\boldsymbol{\Theta}})$ are zero. In addition, the magnitude of this error will be directly affected by the level α for which we are evaluating the SLA-VaR approximation and will asymptotically exact in $\alpha \to 1$ when $e_\alpha(\mathscr{M})$ and $e_\alpha(\widehat{\boldsymbol{\Theta}})$ are zero, otherwise there will be finite errors associated with all three sources of uncertainty in the SLA-VaR.

One readily observes from the simulation study examples provided for these simple LDA models used widely in practice that the SLA as it currently stands may have a relatively large approximation error associated with the estimation of the VaR for non-extreme levels of α. In addition, as the SLA is presented as an asymptotic analytic approximation, it does not present itself readily to uncertainty analysis that arises from the associated uncertainty in the parameter

estimates. We will study this under a simple propagation of error analysis in the following subsection.

8.9.1 UNDERSTANDING THE IMPACT OF PARAMETER ESTIMATION ERROR ON A SLA

It is well known that the estimation of members of the sub-exponential family of statistical models can be difficult to achieve with accuracy in the parameter estimates associated with the parameter representing the heaviness of the tail of the severity distribution. This parameter, however, plays a vital role in the accuracy of the SLA, as detailed in Example 8.3.

Therefore, we note that it is critical that in practical OpRisk settings, the error associated with the estimation of model parameters that parameterize the severity and frequency models in an LDA framework are at a minimum at least considered and reported. This can be achieved typically through the reporting of a confidence interval at a given level of significance associated with the point estimator used to obtain the parameter estimate, as discussed when presenting the definition of $e_\alpha(\widehat{\boldsymbol{\Theta}})$, the error in the SLA associated with parameter estimation error. Such a standard statistical concept is applicable to both frequentist and Bayesian OpRisk modeling paradigms. Furthermore, it is convenient that in several simple examples of severity and frequency models popular in OpRisk, the point estimation of the severity and frequency models can be achieved analytically by a maximum-likelihood-based estimation.

In the previous chapters, we have seen several more advanced models for which this will not be true and one must resort to moment-matching, quantile estimation and numerical maximum-likelihood estimation (MLE)-based parameter estimations. Despite this, it is true that one can obtain confidence intervals associated with the uncertainty in the parameter estimates when fitting the component severity and frequency models in an LDA structure to observed losses and expert opinion.

What is also clear is that, in general, the incorporation of the associated parameter estimation error in the model parameters, as obtained from say a statistical confidence interval, into the estimation of the uncertainty of the VaR-SLA is far from trivial to achieve analytically. In Theorem 8.8, we derive a first-order analytic approximation of the error term $e_\alpha(\widehat{\boldsymbol{\Theta}})$ attributed to the estimation of the SLA-VaR resulting from estimation error in the parameter estimates in the LDA model.

Theorem 8.8 (First-Order SLA-VaR Error w.r.t. Parameter Uncertainty) *Consider an LDA OpRisk model for a given risk type and business cell that has N_t losses in the t-th year satisfying $X_n \overset{\text{i.i.d.}}{\sim} F_X(x;\boldsymbol{\theta}) \in \mathcal{S},\ \forall n \in \{1,\ldots,N_t\}$ and frequency distribution given by an inhomogeneous Poisson process with $N_t \sim Poisson(\lambda_t)$. Furthermore, assume that the estimation of the severity distribution parameter vector $\boldsymbol{\theta}$ is achieve by MLE to produce a random vector point estimator, given observed losses $x_{1:n_t}$ denoted by $\widehat{\boldsymbol{\Theta}}^{\text{MLE}}$. In this case, the SLA-VaR approximation given by*

$$\text{VaR}_\alpha[Z] = F^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right)[1 + o(1)],\ \textit{as}\ \alpha \to 1, \tag{8.202}$$

can be expanded according to a first-order Taylor Series in the LDA model parameters about the point estimators $\widehat{\boldsymbol{\theta}}$ and $\widehat{\lambda}$, denoted by

$$\widehat{\text{VaR}}_\alpha(Z) = F^{-1}\left(1 - \frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1 + o(1)]$$

$$+\nabla_{\boldsymbol{\theta}} F^{-1}\left(1-\frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1+o(1)](\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})$$
$$+\frac{\partial}{\partial\lambda}F^{-1}\left(1-\frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1+o(1)](\lambda-\widehat{\lambda}),$$

as $\alpha \to 1$. *Furthermore, assuming the variance of the parameter point estimators is given by* $\mathbb{C}\text{ov}(\widehat{\boldsymbol{\Theta}}) = \Sigma_{\widehat{\boldsymbol{\Theta}}}$ *and* $\mathbb{V}\text{ar}(\widehat{\lambda}) = \sigma^2_{\widehat{\lambda}}$, *then an expression of first-order error in the SLA-VaR attributed to the parameter uncertainty in the LDA model estimation is given by*

$$\begin{aligned}
\widetilde{e_\alpha}^2(\widehat{\boldsymbol{\Theta}}) &= \mathbb{V}\text{ar}_{\boldsymbol{\Theta},\lambda}(\widetilde{\text{VaR}}_\alpha(Z)) \\
&= \left(\nabla_{\boldsymbol{\theta}} F^{-1}\left(1-\frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1+o(1)]\right) \\
&\quad \Sigma_{\widehat{\boldsymbol{\Theta}}}\left(\nabla_{\boldsymbol{\theta}} F^{-1}\left(1-\frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1+o(1)]\right)^T \\
&\quad +\left(\frac{\partial}{\partial\lambda}F^{-1}\left(1-\frac{1-\alpha}{\lambda};\boldsymbol{\theta}\right)\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}[1+o(1)]\right)^2\sigma^2_{\widehat{\lambda}}.
\end{aligned}$$

Remark 8.22 *The applicability of this first-order approximation is significant because we note that for many popular severity models, analytic expressions for the maximum-likelihood estimator in terms of the data can be obtained analytically. In addition, in some situations, the variance of the parameter estimates may be obtained analytically, in cases in which this is not possible, one can also obtain a lower bound on this error approximation in the SLA-VaR representation because of the parameter estimation via the Cramer–Rao lower bound (CRLB) for the parameter estimation accuracy.*

To illustrate the applicability of Theorem 8.8, we develop expressions for the uncertainty in the SLA-VaR approximation in terms of parameter uncertainty for the three LDA-Pareto models developed. To achieve this, we first provided analytic approximations for the uncertainty in the estimation of the SLA-VaR that can be attributed to an LDA model for which an unbiased, efficient and consistent estimator is obtained analytically for the severity and frequency model parameters. Furthermore, we consider such cases that present a closed-form estimate for the precision in these estimators as a function of the sample size. This is applicable in several examples under MLE for OpRisk LDA models as we will demonstrate.

The first-order approximation of the SLA-VaR error associated with the parameter estimation uncertainty in Poisson-Pareto-LDA models, Poisson-Weibull-LDA models and Poisson-LogNormal-LDA models can be obtained analytically for the error term $e_\alpha(\widehat{\boldsymbol{\Theta}})$. This is obtained according to parameter estimations undertaken via MLE and the resulting error terms are presented in Example 8.11. Each example considers an LDA OpRisk model for a given risk type and business cell that has N_t losses in the tth year satisfying $X_n \overset{\text{i.i.d.}}{\sim} F_X(x;\boldsymbol{\theta}) \in \mathcal{S}$, $\forall n \in \{1,\ldots,N_t\}$ and frequency distribution given by an inhomogeneous Poisson process with $N_t \sim Poisson(\lambda_t)$. Furthermore, assume that the estimation of the severity distribution parameter vector $\boldsymbol{\theta}$ is achieved by MLE to produce a random vector point estimator, given observed losses $x_{1:n_t}$ denoted by $\widehat{\boldsymbol{\Theta}}^{MLE}$.

EXAMPLE 8.11 First-Order SLA for Spectral Risk Measures via CARA Functions

Poisson-Pareto-LDA: Consider the Pareto-LDA model constructed via a severity distribution model in which losses X_i in the tth year are assumed to be distributed according to a two-parameter Pareto distribution in which the shape and scale parameters are restricted to be strictly positive $a > 0$ and $\theta > 0$ given by distribution function

$$F_X(x; a, \theta) = \begin{cases} 1 - \left(\frac{\theta}{x}\right)^a, & x > \theta, \\ 0, & x < \theta. \end{cases} \tag{8.203}$$

Furthermore, we assume that the frequency distribution considered will be Poisson with intensity parameter λ for the mean number of counts $\mathbb{E}[N_t]$ in the t-th year. The maximum-likelihood-estimated parameters after year T, in this model, have a closed-form expression given by

$$\begin{aligned} \widehat{\lambda}^{\text{MLE}} &= \frac{1}{T}\sum_{t=1}^{T} n_t, \\ \widehat{\theta}^{\text{MLE}} &= \min\{x_1, \ldots, x_n\}, \\ \widehat{a}^{\text{MLE}} &= \frac{n}{\sum_{i=1}^{n}(\ln x_i - \ln \widehat{\theta}^{\text{MLE}})}. \end{aligned} \tag{8.204}$$

Under this choice of severity model, one obtains the analytic approximation for the SLA-VaR given by

$$\text{VaR}_\alpha[Z] = F^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\right)[1 + o(1)] = \theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1\right], \ \alpha \to 1.$$

Hence, one obtains the expansion around the parameter estimates given for this model by

$$\begin{aligned} \widetilde{\text{VaR}}_\alpha(Z) = \theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1\right]\Bigg|_{\widehat{\theta},\widehat{\lambda}} + \left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1\right]\Bigg|_{\widehat{\theta},\widehat{\lambda}} (\theta - \widehat{\theta}) \\ - a\theta\left[\left(\frac{\lambda}{1-\alpha}\right)^{-1+1/a}\right]\Bigg|_{\widehat{\theta},\widehat{\lambda}} (a - \widehat{a}) - \frac{a\theta}{1-\alpha}\left[\left(\frac{\lambda}{1-\alpha}\right)^{-1+1/a}\right]\Bigg|_{\widehat{\theta},\widehat{\lambda}} (\lambda - \widehat{\lambda}). \end{aligned}$$

Using this result, one can obtain the resulting error in the SLA-VaR attributed to the parameter uncertainty according to the expression in Theorem 8.8. This requires knowledge of the parameter estimators covariance, which can be obtained for the

Pareto-Poisson-LDA model as

$$\Sigma_{[\widehat{a},\widehat{\theta}]} = \begin{bmatrix} \frac{a}{\theta^2} & -\frac{1}{\theta} \\ -\frac{1}{\theta} & \frac{1}{a^2} \end{bmatrix}^{-1}, \tag{8.205}$$

$$\sigma^2_{\widehat{\lambda}} = \text{Var}\left(\frac{1}{t}\sum_{s=1}^{t} N_s\right) = \frac{1}{t^2}\sum_{s=1}^{t}\text{Var}(N_s) = \frac{\lambda}{t}.$$

In particular, the lower bound on the error associated with the parameter uncertainty in the SLA-VaR can be obtained by substituting the expression for the inverse of the Fisher information matrix ($\Sigma^{CRLB}_{[\widehat{a},\widehat{\theta}]}$) (CRLB) in place of the expression for the variance in the MLE parameter estimates, giving a lower bound for the estimation error given by

$$\widetilde{e}^{\alpha}_2(\widehat{\boldsymbol{\Theta}}) \geq \left(\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1, -a\theta\left(\frac{\lambda}{1-\alpha}\right)^{-1+1/a}\right]\right)\Big|\Sigma^{\text{CRLB}}_{[\widehat{a},\widehat{\theta}]}$$

$$\left(\left[\left(\frac{\lambda}{1-\alpha}\right)^{1/a} - 1, -a\theta\left(\frac{\lambda}{1-\alpha}\right)^{-1+1/a}\right]\right)^T_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}} + \left(\frac{a\theta}{1-\alpha}\left(\frac{\lambda}{1-\alpha}\right)^{-1+1/a}\Big|_{\widehat{\boldsymbol{\theta}},\widehat{\lambda}}\right)^2 \sigma^2_{\widehat{\lambda}}.$$

For more details, see discussions on CRLB and estimators in Cruz *et al.* (2015). ■

In situations in which this first-order approximation is not possible to easily evaluate, an alternative approach to addressing this would involve a numerical analysis. For example, one may consider an alternative Monte-Carlo-based perturbation analysis involving drawing samples of parameter values uniformly from the resulting confidence interval $\widehat{\boldsymbol{\Theta}}^j \sim U[l(\widehat{\boldsymbol{\theta}}, \mathbf{y}, \kappa), u(\widehat{\boldsymbol{\theta}}, \mathbf{y}, \kappa)]$ to obtain a set of J parameter vector realizations $\{\widehat{\boldsymbol{\theta}}^j\}_{j=1:J}$. Then given these parameter estimates, one can evaluate the resulting uncertainty associated with the SLA-VaR approximation attributed to the estimation accuracy of the model parameters, for a given level α and a sample size n_t and confidence κ given by

$$e_{\alpha}(\widehat{\boldsymbol{\Theta}}) \approx \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(F^{-1}\left(1-\frac{1-\alpha}{\lambda_t^j};\boldsymbol{\theta}^j\right) - \frac{1}{J}\sum_{k=1}^{J}F^{-1}\left(1-\frac{1-\alpha}{\lambda_t^k};\boldsymbol{\theta}^k\right)\right)^2}\,[1+o(1)],$$

as $\alpha \to 1$.

Having detailed the analysis of the SLA-VaR approximation accounting for the uncertainty arising from the parameter estimation error in the individual severity and frequency parameter estimates, we now consider the SLA-VaR error term arising from the asymptotic approximation. That is, we study the SLA-VaR approximation error attributed to the SLA approximation component, $e_{\alpha}(SLA)$. We work with the representation developed in Degen (2010) in which this is defined according to the relative approximation error in Definition 8.14.

8.9.2 UNDERSTANDING THE SLA ERROR

We start the discussion on the SLA error for the VaR by defining the relative approximation error as detailed in Definition 8.14.

Definition 8.14 *The relative approximation error in the SLA-VaR error is given by*

$$e_\alpha(\text{SLA}) = \frac{F^{-1}(1-(1-\alpha)/\lambda)}{G^{-1}(\alpha)} - 1,\ \alpha \in (0,1), \tag{8.206}$$

where F represents the severity distribution and G represents the distribution for the annual loss.

In general, this approximation error will be non-analytic and must be approximated numerically. However, under certain conditions on the distribution function for the severity model in the LDA framework, one may be able to assess the relative SLA-VaR approximation error analytically. In particular, as detailed in Omey and Willekens (1986) and Degen (2010), the results on second-order sub-exponentiality are required to proceed with an analytic analysis of the relative error. We will present such results first for sub-exponential family members considered above with asymptotically analytic expressions for the relative SLA-VaR approximation error. Then we present an analytic result for a special family of LDA models based around α-stable models.

In the situations in which one considers heavy-tailed densities which satisfy the condition that they admit regularly varying densities with parameter $-1/\beta - 1$ for some $\beta > 0$, then as $\alpha \to 1$, the result for the representation of the relative approximation error of the SLA-VaR, presented in Lemma 8.4, can be developed.

It will be useful to first provide the following result for the expansion of the expression $(1+x)^n$ given in Lemma 8.3.

Lemma 8.3 *Consider the term $(1+x)^n$, then the following expansion can be shown to hold as $x \to 0$*

$$(1+x)^n = 1 + nx + o(x),\ as\ x \to 0. \tag{8.207}$$

The following results for the representation of the relative error of approximation for the SLA-VaR is significant as it provides means of measuring the accuracy of this annual loss approximation of the tail quantiles. In addition, it was used in Degen (2010) as the main motivation for the refinements for the OpRisk SLA-VaR provided in Definition 8.9.

Lemma 8.4 *Assuming a positive measurable density $f(x)$ for the severity distribution in the LDA model. Furthermore, assume this density satisfies $f \in RV_{-1/\beta-1}$ for some $\beta > 0$ and, therefore, that the tail distribution $\overline{F} \in RV_{-1/\beta}$. Then according to the results presented in Subsection 3.5, which provides the expression of the inverse of the annual loss distribution for tail quantiles given by*

$$G^{-1}(\alpha) = F^{-1}\left(1 - \frac{1-\alpha}{\mathbb{E}[N]}\left(1 + cb\left(F^{-1}(\widetilde{\alpha})\right) + o(b(F^{-1}(\widetilde{\alpha})))\right)^{-1}\right), \tag{8.208}$$

the substitution of this result into the Definition 8.14, Equation 8.206 and using the result in Lemma 8.3, one obtains

$$\begin{aligned}\frac{1}{1+e_\alpha(\text{SLA})} &= \frac{F^{-1}(1-((1-\alpha)/\mathbb{E}[N]))\{1+cb(F^{-1}(\widetilde{\alpha}))+o(b(F^{-1}(\widetilde{\alpha})))\}^{-1}}{F^{-1}(\widetilde{\alpha})} \\ &\sim \frac{(1+cb(F^{-1}(\widetilde{\alpha}))+o(b(F^{-1}(\widetilde{\alpha}))))^{\beta}F^{-1}(\widetilde{\alpha})}{F^{-1}(\widetilde{\alpha})} \\ &= 1+cb(F^{-1}(\widetilde{\alpha}))+o(b(F^{-1}(\widetilde{\alpha}))), \end{aligned} \tag{8.209}$$

with definitions for c *and* $b(\cdot)$ *as discussed in Subsection 3.5. Hence, the relative approximation error can be shown to be expanded asymptotically according to*

$$e_\alpha(\text{SLA}) = -\beta cb(F^{-1}(\widetilde{\alpha}))+o(b(F^{-1}(\widetilde{\alpha}))),\ \textit{as}\ \alpha\to 1. \tag{8.210}$$

Remark 8.23 *This result demonstrates analytically that the SLA-VaR in Definition 8.9 will systematically underestimate the VaR for high levels of* α*, because in the scenarios considered, when* $\beta<2$*, one can show that* $b(\cdot)>0$*,* $c>0$ *and that* c_β *will be monotonically decreasing.*

8.10 Infinite Mean-Tempered Tail Conditional Expectation Risk Measure Approximations

In this section, we discuss a class of risk measures which are special modifications to the ES and TCE risk measures which are a modified class of risk measures that will be categorized as tempered tail conditional expectation (TTCE) risk measures. This class of measures is a special form of a distortion-type risk measure, where the conditional tail expectation which may not exist (i.e. would be infinite) under the standard definition for infinite mean models, will be modified to a risk measure which is a finite version of such a conditional tail expectation. For instance, one could consider the class of fractional lower order moments (FLOMs) which could be utilized as risk measures where the tail expectation is modified according to

$$\text{TTCE}_Z(\kappa) = \mathbb{E}[Z^\alpha | Z > \text{VaR}_\kappa[Z]], \tag{8.211}$$

for some $\alpha\in\mathbb{R}^+$ that makes this expectation finite. When α is not an integer and one sets $\text{VaR}_\kappa[Z]=0$, then this would correspond to the FLOM, see discussion in the context of α-stable models in Chapter 4. This concept could be applied to any classes of heavy-tailed models from EVT, RV models and the resulting FLOMs can be often found in closed form for several families of models.

To illustrate this concept, we will focus on the family of α-stable severity models and their associated closed-form analytic families of LDA models discussed in Chapter 6. In Theorem 8.9, the result for the Poisson-α-stable model is considered, where the mixture weight is obtained after truncation of the total number of losses in a given year. The untruncated case follows trivially as a consequence of such a mixture representation, as discussed in previous chapters on such models. Note that although we start below with the symmetric case, typically we will be more interested in cases where we restrict $\beta=1$ to the perfect skewed setting which will ensure that the support of the severity distribution and resulting annual loss distribution is positive.

Theorem 8.9 (Tail Approximation of LDA Annual Loss Truncated-Poisson-α-Stable) *Given the number of losses is distributed $N \sim Poisson(\lambda)\mathbb{I}[N \leq N_{\max}]$ and for all $i \in \{1, \ldots, N\}$ each loss random variable is distributed as α-stable under the $S(0)$ parametrization, then the resulting right tail of the density of the compound process (annual loss under an LDA framework) is given as $z \to \infty$ by*

$$f_Z(z) \sim \sum_{n=1}^{N_{\max}} \frac{N_{\max}!}{\Gamma(N_{\max}+1,\lambda)} \exp(-\lambda)\frac{\lambda^n}{n!}\alpha\widetilde{\gamma}_n^{\alpha}c_{\alpha}(1+\widetilde{\beta}_n)z^{-(\alpha+1)}, \tag{8.212}$$

where $\Gamma(N_{\max}+1,\lambda)$ is the incomplete gamma function and

$$c_{\alpha} = \sin\left(\frac{\pi\alpha}{2}\right)\frac{\Gamma(\alpha)}{\pi}.$$

Furthermore, the upper right tail of the annual loss distribution is given as $z \to \infty$ by

$$\begin{aligned}\bar{F}_Z(z) \sim \sum_{n=1}^{N_{\max}} \frac{N_{\max}!}{\Gamma(N_{\max}+1,\lambda)} \exp(-\lambda)\frac{\lambda^n}{n!}\widetilde{\gamma}_n^{\alpha}c_{\alpha}(1+\widetilde{\beta}_n)z^{-\alpha}\mathbb{I}[\widetilde{\delta}_n \leq z \leq N_{\max}] \\ + \exp(-\lambda)\mathbb{I}[z=0].\end{aligned} \tag{8.213}$$

Remark 8.24 *As discussed in Chapter 4, as a consequence of the reflection property of the α-stable severity model, one can also characterize the lower tail properties for admissible parameter values satisfying $\alpha < 2$ and $-1 < \beta < 1$. This result extends to the compound Poisson tails trivially as above.*

Having detailed the tail properties of the annual loss distribution and density functions, we next consider derivation of other analytic properties for this class of LDA models. In particular, we start with the expected loss and progress to expected values of functions of the annual loss. Finishing with estimations of VaR and ES for these LDA α-stable models.

As detailed previously the expected loss for such severity models and, therefore, for the resulting annual loss model is known analytically according to the results in Theorem 8.10. This uses the fact that for expected powers of an α-stable random variable will be finite according to

$$\mathbb{E}[X^p] < \infty \quad \text{if } -1 < p < \alpha, \text{ and } \alpha \in (0,2]. \tag{8.214}$$

Note that not only can one demonstrate existence (finite values) of the FLOMs but also, in addition, for two different cases, one can find analytic expressions for such moments with respect to the model parameters.

1. Firstly, if $\alpha > 1$ and $p = 1$, then the mean is analytically known for the severity model.
2. Secondly, when a symmetric α-stable model is appropriate for the severity model, for example, consider modeling cases under either *Scenario 2*, *Scenario 3* or *Scenario 4* in Chapter 4, then one can also provide an analytic expression for the FLOMs in terms of the parameters.

For a generic symmetric α-stable model, the following analytic results are known for the FLOM according to Lemma 8.5, see details in Chapter 4 and 2011.

Lemma 8.5 (α-Stable Severity Fractional Lower Order Moments (FLOMs)) *Given a sub-exponential α-stable severity distribution with losses $X \sim S_\alpha(x; \beta, \gamma, \delta; S(0))$, then the expected loss can be evaluated analytically according to the following moments or FLOM conditions for $p \in (-1, 2)$,*

$$\mathbb{E}[|X|^p] = \begin{cases} +\infty, & p > \alpha, \\ 2\Gamma(1+p)\sin(\frac{\pi}{2}p) & \\ \quad \int_0^\infty (1 - \mathscr{R}e\{\Phi(t)\})t^{p-1}dt, & 0 < p < \alpha < 1, \beta \neq 0, \\ C(p, \alpha; S(0))\gamma^{\frac{p}{\alpha}}, & -1 < p < \alpha < 2, \beta = 0, \end{cases} \tag{8.215}$$

where $\Phi(t)$ represents the characteristic function of loss r.v. X and

$$C(p, \alpha) = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\Gamma\left(-\frac{p}{2}\right)},$$

and the following special cases of the first-integer moment given by

$$\mathbb{E}[X] = \begin{cases} \delta - \beta\gamma\tan\frac{\pi\alpha}{2}, & p = 1, 1 < \alpha < 2, \\ \delta, & p = 1, \alpha = 2. \end{cases} \tag{8.216}$$

The things to observe from these results are that when $\beta = 0$, then for $\alpha > 1$, the mean is at δ and the mode is at δ. Therefore, if the location parameter is large as indicative of a heavy-tailed severity model which may have very large mean, the tail probability of a negative loss $\mathbb{P}\mathrm{r}[X < 0] \approx 0$, therefore, can be considered as negligible, making Scenario 3 from Chapter 6 a suitable approximation. Alternatively, when modeling the log-transformed data, it is clearly going to prove convenient to restrict to model cases in which the skewness parameter is symmetric $\beta = 0$, should one wish to evaluate the expected annual loss or the expectation of functions of the annual loss for quantify things such as ES. The reason for this is that the representation of the FLOM expressions makes analytic results achievable for models with strictly stable severity distributions.

An alternative representation of the FLOMs results are derived for the parametrization $S(1)$ in Samorodnitsky and Taqqu (1994a) and gives the results for the FLOMs according to Lemma 8.6.

Lemma 8.6 *Given a sub-exponential α-stable severity distribution with losses $X \sim S_\alpha(x; \beta, \gamma, \delta; S(1))$ in which $\alpha \in (0, 2)$ and $\beta = 0$ when $\alpha = 1$. Then the expected loss can be evaluated analytically according to the following moments or FLOM conditions for $p \in (0, \alpha)$ according to*

$$\mathbb{E}[|X|^p] = C(p, \alpha, \beta; S(1))^p\gamma, \tag{8.217}$$

where Hardin Jr (1984) and Samorodnitsky and Taqqu (1994a, property 1.2.17) showed that

$$C(p,\alpha,\beta;S(1)) = \frac{2^{p-1}\Gamma\left(1-\frac{p}{\alpha}\right)}{p\int_0^\infty u^{-p-1}\sin^2 u du}\left(1+\beta^2\tan^2\frac{\alpha\pi}{2}\right)^{\frac{p}{2\alpha}} \times \cos\left(\frac{p}{\alpha}\arctan\left(\beta\tan\frac{\alpha\pi}{2}\right)\right). \tag{8.218}$$

Given these expressions for the fractional moments, one can derive the following expected annual loss results given in Theorem 8.10 which will be beneficial for many loss approximation methods, which include refinement of the SLA for risk measures such as VaR and, in particular, in estimation of the ES.

Theorem 8.10 (LDA-Expected Annual Loss Poisson-α-Stable) *Given the number of losses is distributed $N \sim Poisson(\lambda)$ and for all $i \in \{1, \ldots, N\}$ and each loss random variable is distributed $X_i \sim S_\alpha(x;\beta_i,\gamma_i,\delta_i;S(0))$, then the expected annual loss under an LDA compound process framework is given for $1 < \alpha \leq 2$ by*

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{n=1}^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}\int_{-\infty}^{\infty} z\, dS_\alpha(x;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;0) \\ &= \sum_{n=1}^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}\left(\widetilde{\delta}_n - \widetilde{\beta}_n\widetilde{\gamma}_n\tan\left(\frac{\pi\alpha}{2}\right)\right).\end{aligned} \tag{8.219}$$

As a consequence of this result for the annual expected loss for this class of stable distributions, one can determine as a function of the severity model parameters whether quantities such as tail expectations will exist.

Corollary 8.2 *The results in Theorem 8.10 demonstrate that the expected loss is infinite for models in which the estimated or expert elicited value of the tail index α is found to be less than 1. This also results in the ES being infinite for any such model, whereas the VaR in such cases will always be finite.*

Remark 8.25 *One can also observe that although the conditional expectations with respect to the tail of the annual loss process may be infinite for such annual loss processes for $\alpha < 1$, there is still significant interest in moments of functions of the annual loss process, which may be finite.*

Clearly, it will depend on the function specified as to the existence of such moments. In general, such processes are directly of interest to insurance modeling in OpRisk. For example, piecewise linear functions of the annual loss, known as haircuts, may be applied as a result of certain insurance contracts aimed at mitigating the risk associated with potential catastrophic losses. Furthermore, one may be working under the modeling framework proposed for limited liability annual loss LDA models as presented in Theorem 6.10, in such cases, the calculation of moments is critical.

Remark 8.26 *To illustrate the applicability of these ideas beyond just the α-stable model setting, we observe that the FLOM for the Pareto model can also be obtained for each of the different types*

of Pareto model Type I–IV (see the EVT Chapter 2). The Pareto distribution Type I–IV can be categorized as follows according to the tail of the distribution:

$$\overline{F}(x;a,\theta;I)=\left(\frac{x}{\theta}\right)^{-a},\ x>\theta,\ \theta>0,a\in\mathbb{R},$$
$$\overline{F}(x;a,\theta,\mu;II)=\left(1+\frac{x-\mu}{\theta}\right)^{-a},\ x>\mu,\ \theta>0,a\in\mathbb{R},\mu\in\mathbb{R},$$

$$\overline{F}(x;a,\theta;\text{Lomax})=\left(1+\frac{x}{\theta}\right)^{-a},\ x>0,\ \theta>0,a\in\mathbb{R}, \tag{8.220}$$
$$\overline{F}(x;\theta,\mu,\gamma;III)=\left(1+\left(\frac{x-\mu}{\theta}\right)^{\frac{1}{\gamma}}\right)^{-1},\ x>\mu,\ \mu\in\mathbb{R},\theta>0,\gamma>0,$$
$$\overline{F}(x;a,\theta,\mu,\gamma;IV)=\left(1+\left(\frac{x-\mu}{\theta}\right)^{\frac{1}{\gamma}}\right)^{-a},\ x>\mu,\ \mu\in\mathbb{R},\theta>0,\gamma>0,a\in\mathbb{R}.$$

These sub-families of the Pareto family of distribution types will then produce FLOMs given by

$$\begin{aligned}
\mathbb{E}[X^p;I]&=\frac{\theta^p a}{a-p}, && p<a,\\
\mathbb{E}[X^p;II]&=\frac{\theta^p\Gamma(a-p)\Gamma(1+p)}{\Gamma(a)}, && p\in(-1,a),\\
\mathbb{E}[X^p;III]&=\theta^p\Gamma(1-\gamma p)\Gamma(1+\gamma p), && p\in\left(-\frac{1}{\gamma},\frac{1}{\gamma}\right),\\
\mathbb{E}[X^p;IV]&=\frac{\theta^p\Gamma(1-\gamma p)\Gamma(1+\gamma p)}{\Gamma(a)}, && p\in\left(-\frac{1}{\gamma},\frac{a}{\gamma}\right).
\end{aligned} \tag{8.221}$$

Returning back to the α-stable examples, next we discuss how given an annual loss random variable Z distributed according to a mixture of α-stable distributions, as presented in Theorem 6.8, one can still obtain analytic results in certain cases for moments of functions of the annual loss Z for which moments may exist. Functions of the annual loss random variable can arise in several scenarios, including evaluation of corrections to SLAs for VaR, ES and, of course, SRM. In addition, one can conceive of such instances involving evaluation of the expected loss in the presence of insurance haircuts and transformations, and evaluation of tail expectations such as for evaluation of risk measures such as ES. Under such functions of the annual loss random variable, one can utilize the following results to obtain the expected loss analytically, or alternatively the expected tail loss.

Consider a function $g:\mathbb{R}\to\mathbb{R}$ which is a function of the annual loss $g(Z)$. Assume one is interested in calculating expectations of functions of such an annual loss random variables, such as

$$\mathbb{E}[g(Z)]=\sum_{n=1}^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}\int_0^{\infty}g(z)dS_{\alpha}(x;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0)). \tag{8.222}$$

For example, one may be interested in functions of the annual loss such as

$$g(Z) = (Z-a)^p, \text{ for some constant } a \in \mathbb{R}, p \in \mathbb{J};$$
$$g(Z) = \ln(Z);$$
$$g(Z) = \exp(Z);$$
$$g(Z) = H(Z-a) = \begin{cases} 1, \text{if } Z \geq a, \\ 0, \text{otherwise;} \end{cases}$$
$$g(Z) = ZH(Z-a) = \begin{cases} Z, \text{if } Z \geq a, \\ 0, \text{otherwise;} \end{cases}$$
$$g(Z) = Z^p H(Z-a) = \begin{cases} Z^p, \text{if } Z \geq a, \\ 0, \text{otherwise;} \end{cases}$$
$$g(Z) = w(h(z)), \text{for some composite functions } w(\cdot), h(\cdot).$$

It is clear that to obtain analytic expressions of expectations of such random variates in the case of the LDA Poisson-stable model; one would not be able to perform standard approaches such as taking Taylor and Maclaurin series expansions in integer powers of polynomial terms and then evaluating the moment-generating function for such powers to estimate these expectations of polynomial terms. This is complicated by two related facts: firstly, the moment-generating function is non-existent at the origin for α-stable models with $\alpha < 1$ and secondly, the moments of integer orders are not finite for powers exceeding 1 for $\alpha \in (1, 2]$ and no integer powers for $\alpha \in [0, 1]$.

In the following section, we will illustrate how calculation of expectations of such functionals will result in *exact* analytic solutions for the ES for the LDA α-stable models. To achieve this, we recall the definition of the ES provided in Cruz *et al.* (2015, chapter 6) in the context of generic LDA α-stable models.

Theorem 8.11 (Closed-Form ES for Generic LDA Models: α-Stable Severity) *Consider the number of losses $N \sim F_N(n)$ with $N \in \{0, 1, \ldots, N_{\max}\}$ and $N_{\max}$ can be infinite, where for all $i \in \{1, \ldots, N\}$, each loss random variable is distributed as α-stable. If one considers the general LDA $F_N(n)$-α-Stable model in which modeling the severity process with a finite mean is sensible, thus restricting $\alpha \in (1, 2]$, then the ES reduces to the tail expectation for all quantiles exceeding $\mathbb{P}\mathrm{r}[Z = 0]$, after which the density of the annual loss process is provably continuous for the class of LDA α-stable models. Therefore, in terms of a function of the annual loss, $g(Z)$, one may express the ES according to*

$$\begin{aligned} &\mathbb{E}[Z \mid \{Z > \mathrm{VaR}_\alpha[Z]\}] \\ &= \sum_{n=1}^{N_{\max}} \mathbb{P}\mathrm{r}[N = n] c(\alpha, \beta, \gamma, \delta, \mathrm{VaR}_\alpha[Z]) \mathbb{E}_{S_\alpha(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S(0))}[g(Z)] \\ &= \sum_{n=1}^{\infty} \mathbb{P}\mathrm{r}[N = n] c(\alpha, \beta, \gamma, \delta, \mathrm{VaR}_\alpha[Z]) \int_{-\infty}^{\infty} g(Z) dS_\alpha(z; \widetilde{\beta}_n, \widetilde{\gamma}_n, \widetilde{\delta}_n; S(0)), \end{aligned} \tag{8.223}$$

where one selects the Heaviside function according to

$$g(Z) = ZH(Z - \mathrm{VaR}_\alpha[Z]) = \begin{cases} Z, & \text{if } Z \geq \mathrm{VaR}_\alpha[Z], \\ 0, & \text{otherwise;} \end{cases} \tag{8.224}$$

for all levels $\alpha \in (0,1)$ *such that* $\text{VaR}_\alpha[Z] > \mathbb{P}\text{r}[N=0]$ *and where*

$$c(\alpha,\beta,\gamma,\delta,\text{VaR}_\alpha[Z]) =$$
$$\begin{cases} \left(-\frac{1}{\pi\alpha}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(n\alpha+1)}{n\Gamma(n+1)}\sin(\pi n\rho)\left(\frac{\text{VaR}_\alpha[Z]-\delta}{\gamma^\alpha}\right)^{-\alpha n}\right)^{-1}, & \text{if } \alpha<1, \\ \left(\frac{1}{2}\left(1+\beta\frac{K(\alpha)}{\alpha}\right)\right. & \\ \quad \left.+\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma\left(\frac{n}{\alpha}+1\right)}{n\Gamma(n+1)}\sin(\pi n\rho)\left(\frac{\text{VaR}_\alpha[Z]-\delta}{\gamma^\alpha}\right)^{n}\right)^{-1}, & \text{if } \alpha>1, \\ \left(-\frac{1}{\pi}b_0+\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}b_n\left(\frac{\text{VaR}_\alpha[Z]-\delta}{\gamma}\right)^{n}\right)^{-1}, & \alpha=1,\beta>0. \end{cases}$$

To illustrate this with two examples, consider the definitions for the exact solutions to the Poisson example of closed-form expressions for ES given as follows.

EXAMPLE 8.12 Exact Closed-Form Expected Shortfall for LDA Poisson-α-Stable

Given the number of losses is distributed as $N \sim Poisson(\lambda)$ and for all $i \in \{1,\ldots,N\}$ with each loss random variable X_i is distributed as α-stable. If one considers the general LDA Poisson-α-stable model in which modeling the severity process with a finite mean is sensible, thus restricting $\alpha \in (1,2]$, then the ES reduces to the tail expectation for all quantiles exceeding $\mathbb{P}\text{r}[Z=0] = \exp(-\lambda)$, after which the density of the annual loss process is provably continuous for the class of LDA α-stable models. Therefore, in terms of a function of the annual loss, $g(Z)$, one may express the ES according to

$$\begin{aligned} &\mathbb{E}[Z \mid \{Z > \text{VaR}_\alpha[Z]\}] \\ &= \sum_{n=1}^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}c(\alpha,\beta,\gamma,\delta,\text{VaR}_\alpha[Z])\mathbb{E}_{S_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0))}[g(Z)] \qquad (8.225) \\ &= \sum_{n=1}^{\infty}\exp(-\lambda)\frac{\lambda^n}{n!}c(\alpha,\beta,\gamma,\delta,\text{VaR}_\alpha[Z])\int_{-\infty}^{\infty}g(Z)dS_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0)), \end{aligned}$$

where one selects the Heaviside function according to

$$g(Z) = ZH(Z-\text{VaR}_\alpha[Z]) = \begin{cases} Z, & \text{if } Z \geq \text{VaR}_\alpha[Z], \\ 0, & \text{otherwise;} \end{cases} \qquad (8.226)$$

for all levels $\alpha \in (0,1)$ such that $\text{VaR}_\alpha[Z] > \exp(-\lambda)$ and where

$$c(\alpha,\beta,\gamma,\delta,\mathrm{VaR}_\alpha[Z]) = \begin{cases} \left(-\frac{1}{\pi\alpha}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma(n\alpha+1)}{n\Gamma(n+1)}\sin(\pi n\rho)\left(\frac{\mathrm{VaR}_\alpha[Z]-\delta}{\gamma^\alpha}\right)^{-\alpha n}\right)^{-1}, & \text{if } \alpha<1, \\ \left(\frac{1}{2}\left(1+\beta\frac{K(\alpha)}{\alpha}\right)\right. & \\ \quad \left.+\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}\frac{\Gamma\left(\frac{n}{\alpha}+1\right)}{n\Gamma(n+1)}\sin(\pi n\rho)\left(\frac{\mathrm{VaR}_\alpha[Z]-\delta}{\gamma^\alpha}\right)^{n}\right)^{-1}, & \text{if } \alpha>1, \\ \left(-\frac{1}{\pi}b_0+\frac{1}{\pi}\sum_{n=1}^{\infty}(-1)^{n-1}b_n\left(\frac{\mathrm{VaR}_\alpha[Z]-\delta}{\gamma}\right)^{n}\right)^{-1}, & \alpha=1,\beta>0. \end{cases} \tag{8.227}$$

■

In general, such definitions can also be obtained for the binomial, negative binomial and doubly stochastic versions of these LDA models, and in all cases, the integral given by

$$\mathbb{E}_{S_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0))}[g(Z)] = \int_{-\infty}^{\infty} g(Z)dS_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0)), \tag{8.228}$$

will be present in different linear-weighted linear combinations, depending on which of the models is considered (Poisson, truncated Poisson, binomial, negative binomial and their doubly stochastic counterparts). Therefore, the main aim of obtaining closed-form expressions for the risk measures such as ES for such models becomes finding a solution to such expectations with respect to particular choice of function g. Of course, when written in this generic notation, it can also capture the SRM for this class of models as discussed earlier.

To proceed with a general framework for approximation of this expectation in Equation 8.228 that will allow approximation of the class of TTCE risk measures, for ultra-heavy-tailed severity models for which the first moment may not be finite, we first note the following property in the α-stable case given in Lemma 8.7, see Zolotarev (1983).

Lemma 8.7 *Given a random variable X with stable law $F_X(x) = S_\alpha(x;\beta,\gamma,\delta;S(0))$, then without loss of generality considering the density $f_X(x) = S_\alpha(x;\beta,1,0;S(0))$ for all admissible parameters the derivative of any order $n \in \mathbb{J}^+$ is bounded above by*

$$\left|\frac{d^n}{dx^n}\, f_X(x)\right| \leq \frac{1}{\pi\alpha}\Gamma\left(\frac{n+1}{\alpha}\right)\left(\cos\left[\frac{\pi}{2}K(\alpha)\beta\right]\right)^{-\frac{n+1}{\alpha}}. \tag{8.229}$$

Therefore, the α-stable laws have densities with uniformly bounded derivatives of any order and as a consequence $f_X(x) \in \mathrm{C}^n(\mathbb{R})$ for all n. Hence, the densities are continuous for all elements of $x \in \mathbb{R}$.

This differentiability result will be directly applicable to the series-expansion-based approach that will be considered in approximation of the integral in Equation 8.228. To develop this approach, we first need to present a few additional concepts based on fractional

calculus. It turns out that a little known result on Taylor series and Maclaurin series expansions of fractional powers discussed in the papers of Usero (2007), Munkhammar (2005), Lavoie *et al.* (1976), Podlubny (1999) and Oldham and Spanier (1974) can be utilized to combine with the analytic results presented in Theorem 6.8 to obtain analytic expectations, provided certain conditions are satisfied as detailed below.

The first step is to decide the appropriate form of fractional derivative as they are not unique; there are choices ranging from the original Riemann–Liouville form Feller (1966) and Grunwal–Letnikov specifications. In the first two chapters of Ortigueira (2011), there is a detailed and informed discussion on the differences and relationships between different fractional derivatives that have been defined over the past 300 years, and we note that there is still an ongoing debate being studied as to the most appropriate form of fractional derivative. We do not enter into this discussion, instead we utilize one of the most widely known forms of fractional derivative utilized in physics, engineering and increasingly stochastic analysis.

In Theorem 8.16, a presentation of fractional Taylor series expansions with respect to Caputo specifications are developed in two parts, first for basic functions of Z and then for composite functions. These fractional series expansions of such general functions can then be truncated at a desired order of approximation and the expectation of these fractional series can be obtained with respect to the heavy-tailed annual loss distribution model to obtain an approximation to the TTCE risk measure.

In order to create such fractional series expansions for composite functions, it will be advantageous to recall a few basic theorems from calculus and their fractional counterparts. We start by defining the choice of fractional derivative that will be considered in this chapter, the Caputo fraction derivative (CFD) denoted by upper script D^C and given in Definition 8.15.

Definition 8.15 (Caputo Fractional Derivative) *Let $n = [\zeta] + 1$ for $\zeta \in (0, 1)$, then the right and left CFDs are given, respectively, by*

$$D^{C,\zeta}_{a|x} f(x) = I^{n-\zeta}_{a|x} \frac{d^n}{dx^n} f(x) = \frac{1}{\Gamma(n-\zeta)} \int_a^x (x-\gamma)^{n-\zeta-1} \frac{d^n}{d\gamma^n} f(\gamma) d\gamma,$$

$$D^{C,\zeta}_{x|b} f(x) = I^{n-\zeta}_{x|b} \left(-\frac{d}{dx}\right)^n f(x) = \frac{1}{\Gamma(n-\zeta)} \int_x^b (\gamma-x)^{n-\zeta-1} \left(-\frac{d}{d\gamma}\right)^n f(\gamma) d\gamma,$$

where $I^{n-\zeta}_{a|x}$ and $I^{n-\zeta}_{x|b}$ denote the right and left Riemann–Liouville fractional integrals, respectively.

Note that as discussed in Baleanu and Trujillo (2010) the fractional derivatives have the property that they are the infinitesimal generators of a class of translation invariant convolution semigroups appearing universally as attractors. It is also worth noting that in order to proceed, one should be aware that the forms of the fractional Leibnitz rule, the fractional chain rule and the fractional Taylor series differ to classical definitions. Here we provide the details regarding the fractional versions of these fundamental results in calculus.

First, we note that as a consequence of the above CFD definition, one has the following results for derivatives of example functions of relevance in many applications.

$$D^{C,\zeta}_{a|x} (x-a)^\beta = \frac{\Gamma(\beta+1)}{\Gamma(\beta-\zeta+1)} (x-a)^{\beta-\zeta}, \ \beta \neq 0, 1, \ldots, [\zeta];$$

$$
\begin{aligned}
D_{x|b}^{C,\zeta}(b-x)^{\beta} &= \frac{\Gamma(\beta+1)}{\Gamma(\beta-\zeta+1)}(b-x)^{\beta-\zeta},\ \beta \neq 0,1,\ldots,[\zeta];\\
I_{a|x}^{\zeta}(x-a)^{\beta} &= \frac{\Gamma(\beta+1)}{\Gamma(\beta+\zeta+1)}(x-a)^{\beta+\zeta};\\
I_{x|b}^{\zeta}(b-x)^{\beta} &= \frac{\Gamma(\beta+1)}{\Gamma(\beta+\zeta+1)}(b-x)^{\beta+\zeta}.
\end{aligned}
\tag{8.230}
$$

In addition, one can show that the CFD of a constant is zero from the left- and right-sided definitions. Note that this is not true for other forms of fractional derivative such as the Riemann–Liouville choice. Having defined the CFD, next we recall the fundamental theorem of calculus (FTC) in Theorem 8.12 and the fractional calculus equivalent for the case of the CFD.

Theorem 8.12 (Fundamental Theorem of Calculus) *Consider a function $f(x)$, then the following derivative and integral operators are inverses of each other producing*

$$
\begin{aligned}
&\textstyle\int_a^b dx \frac{d}{dx} f(x) = f(b) - f(a),\\
&\textstyle\frac{d}{dx}\int_a^x f(s)ds = f(s).
\end{aligned}
\tag{8.231}
$$

Hence, the derivative is the inverse of the integral and vice versa.

The CFD analogue of the FTC known here as the fractional fundamental theorem of calculus (FFTC) is given in Theorem 8.13, see Li *et al.* (2009) and Miller and Ross (1993).

Theorem 8.13 (Fractional Fundamental Theorem of Calculus) *Consider a function $f(x)$, then the following fractional derivative and integral operators are the fractional inverses of each other producing one of the following two results:*

1. *If one considers a number $\zeta > 0$ and a function $f(x) \in L_{\infty}(a,b)$ or $f(x) \in C[a,b]$, then one has*

$$
\begin{aligned}
D_{a|x}^{C,\zeta} I_{a|x}^{\zeta} f(x) &= f(x),\\
D_{x|b}^{C,\zeta} I_{x|b}^{\zeta} f(x) &= f(x).
\end{aligned}
\tag{8.232}
$$

   Here $L_{\infty}(a,b)$ is the set of Lebesgue measurable functions on (a,b) for which $\|f\|_{\infty} < \infty$ and $AC[a,b]$ represents the space of functions f which is absolutely continuous on interval $[a,b]$ and finally $C[a,b]$ is the space of continuous functions f on $[a,b]$ with norm $\|f\|_C = \max_{x\in[a,b]}|f(x)|$. Hence, the derivative is the inverse of the integral and vice versa.
2. *If one considers a number $\zeta \in (0,1)$ and $f(x) \in AC[a,b]$ or $f(x) \in C[a,b]$, then one has*

$$
\begin{aligned}
I_{a|x}^{\zeta} D_{a|x}^{C,\zeta} f(x) &= f(x) - f(a),\\
I_{x|b}^{\zeta} D_{x|b}^{C,\zeta} f(x) &= f(x) - f(b).
\end{aligned}
\tag{8.233}
$$

It will also be relevant in the build up to the results for Fractional Taylor series expansions, to also present the Faa di Bruno theorem for differentiation of composite functions and the corresponding fractional form. The basic definition for the generalized chain rule otherwise

known as the Faa di Bruno theorem is given in Theorem 8.14; see Abramowitz and Stegun (1965, p. 823 Faa di Bruno's Formula).

Theorem 8.14 (Faa di Bruno's Formula) *The generalized chain rule for integer powers of a composite function $f(g(x))$ is given by*

$$D_x^n f\left(g(x)\right) = \sum \frac{n!}{m_1!1!^{m_1} m_2!2!^{m_2} \cdots m_n!n!^{m_n}} \underset{g(x)}{\overset{m_1+\cdots+m_n}{\mathrm{D}}} f(g(x)) \prod_{j=1}^{n} (D_x^j g(x))^{m_j}, \tag{8.234}$$

where $D_{g(x)}^n f(x)$ denotes the nth derivative of $f(x)$ with respect to function $g(x)$ and the summation is overall n-tuples of non-negative integers $(m_1, \ldots, m_n)$ satisfying the constraint that

$$m_1 + 2m_2 + 3m_3 + \cdots + nm_n = n.$$

An alternative way to write this generalization of the chain rule for composite functions is to write it in terms of Bell polynomials according to

$$D_x^n f(g(x)) = \sum_{k=0}^{n} D_{g(x)}^k f(g(x)) B_{n,k}(D_x g(x), D_x^2 g(x), \ldots, D_x^{n-k+1} g(x)). \tag{8.235}$$

Having defined the general expression for the composite function derivatives under standard differentiation, one can derive the CFD equivalent fractional Faa di Bruno theorem given in Theorem 8.15, see Baleanu and Trujillo (2010).

Theorem 8.15 (Fractional Derivatives of Composite Functions via Caputo Derivatives) *Consider analytic composite function $\phi(x) = w(h(x))$, then the left CFD of power ζ for the composite function is given by*

$$\begin{aligned} D_{a|x}^{C,\zeta} w(h(x)) = {} & \frac{(x-a)^{-\zeta}}{\Gamma(1-\zeta)} (w(h(x)) - w(h(a))) \\ & + \sum_{m=0}^{\infty} C_m^{\zeta} m! \frac{(x-a)^{m-\zeta}}{\Gamma(m-\zeta+1)} \sum_{k=1}^{m} D_{h(x)}^m w(h(x)) \\ & \sum_{a_r>0} \prod_{r=1}^{m} \frac{1}{a_r!} \left(\frac{D_x^r h(x)}{r!} \right)^{a_r}, \end{aligned} \tag{8.236}$$

where the sum $\sum_{a_r>0}$ is understood to be overall combinations of non-negative values of $a_1, \ldots, a_m$ such that $\sum_{r=1}^{m} r a_r = m$ and $\sum_{r=1}^{m} a_r = k$ with a_r the Faa di Bruno coefficients defined in Definition 8.15.

From these results and definitions, we can now proceed to define the fractional Taylor series expansions in Theorem 8.16, see Li *et al.* (2009) and Usero (2007, theorem 2.1).

Theorem 8.16 (Caputo-Fractional-Derivative-Based Fractional Taylor Series) *Let $\zeta \in (0, 1]$, $n \in \mathbb{J}$ and $f(x)$ a continuous function in $[a, b]$ satisfying the following conditions:*

1. *$\forall m = 1, \ldots, n$, one has that $(D_{a|x}^{C,\zeta})^m f(x) \in \mathbb{C}([a, b])$ and $(D_{a|x}^{C,\zeta})^m f(x) \in \mathscr{J}_{a,\zeta}$ where $\mathscr{J}_{a,\zeta} = \{f \in \mathbb{C}(\Omega) : I_a^{\gamma} f(x)$ exists and is finite in $\Omega\}$.*

2. *$(D^{C,\zeta}_{a|x})^m f(x)$ is continuous on $[a,b]$.*

Then $\forall x \in [a,b]$, the general series expansion on right or left can be obtained as follows:

- ***Right Series.*** *The right series expansion is given by assuming that $f(x)$ is a function defined on the right neighbourhood of a and that it is an infinitely fractionally differentiable function at a, such that $(D^{\zeta}_{a|x})^m f(x)$ for $m \in \mathbb{J}$, with $D^{\zeta}_{a|x}\cdot$ applied m-times such that $(D^{\zeta}_{a|x})^m = D^{\zeta}_{a|x} D^{\zeta}_{a|x} \cdots D^{\zeta}_{a|x}$. Then the fractional right-Riemann–Liouville Taylor series of function $f(x)$ is given by*

$$f(x) = \sum_{m=0}^{\infty} (D^{\zeta}_{a|x})^m f(x)|_{x=a}[(I^{\zeta}_{a|x})^m \cdot 1], \tag{8.237}$$

where

$$[(I^{\zeta}_{a|x})^m \cdot 1] = \frac{1}{\Gamma(m\zeta+1)}(x-a)^{m\zeta}. \tag{8.238}$$

- ***Left Series.*** *The left series expansion is given by assuming that $f(x)$ is a function defined on the left neighbourhood of b and that it is an infinitely fractionally differentiable function at b, such that $(D^{\zeta}_{x|b})^m f(x)$ for $m \in \mathbb{J}$, with $D^{\zeta}_{x|b}\cdot$ applied m-times such that $(D^{\zeta}_{x|b})^m = D^{\zeta}_{x|b} D^{\zeta}_{x|b} \cdots D^{\zeta}_{x|b}$. Then the fractional right-Riemann–Liouville Taylor series of function $f(x)$ is given by*

$$f(x) = \sum_{m=0}^{\infty} (D^{\zeta}_{x|b})^m f(x)|_{x=b}[(I^{\zeta}_{x|b})^m \cdot 1], \tag{8.239}$$

where

$$[(I^{\zeta}_{x|b})^m \cdot 1] = \frac{1}{\Gamma(m\zeta+1)}(b-x)^{m\zeta}. \tag{8.240}$$

Remark 8.27 *Hence, one has a way of performing the series expansion from either end initial point a or b. An alternative representation in terms of remainder terms is given by Usero (2007, theorem 2.1) where the right series can be represented according to the first n fractional derivatives that may exist, according to Equation 8.241.*

$$f(x) = \sum_{m=0}^{n} (D^{\zeta}_{a|x})^m f(x)|_{x=a} \cdot [(I^{\zeta}_{a|x})^m \cdot 1] + R_n(x,a), \tag{8.241}$$

where the remainder term is given by

$$R_n(x,a) = (D^{\zeta}_{a|x})^{n+1} f(x)|_{x=s} \cdot [(I^{\zeta}_{a|x})^{n+1} \cdot 1], \quad a \le s \le x. \tag{8.242}$$

The validity of these result is predicated on the development recently of the pre-requisite that makes fractional series expansion possible, given by the fundamental theorem of fractional calculus, which is only recently proved and highlighted in Kilbas et al. *(2006) and Tarasov (2008).*

Returning back to the integral expression of interest in the TTCE proposed,

$$\mathbb{E}_{S_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0))}[g(Z)] = \int_{-\infty}^{\infty} g(Z) dS_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0)), \tag{8.243}$$

then for suitable CFD differentiable functions $g(z)$, one can approximate the above expectation by the following steps:

1. First, take, for instance, the right series expansion in Theorem 8.16 of the function $g(z)$ for a suitable point a that could depend on the choice of function $g(z)$ used or on a simplification that makes the future expectations of this function tractable and also note that

$$[(I^{\zeta}_{a|z})^m \cdot 1] = \frac{1}{\Gamma(m\zeta+1)}(z-a)^{m\zeta}. \tag{8.244}$$

One then obtains the function approximation given by nth order truncation according to

$$\begin{aligned}\widehat{g}(z) &= \sum_{m=0}^{n} (D^{\zeta}_{a|z})^m g(z)|_{z=a}[(I^{\zeta}_{a|z})^m \cdot 1] \\ &= \sum_{m=0}^{n} \frac{1}{\Gamma(m\zeta+1)}(D^{\zeta}_{a|z})^m g(z)|_{z=a}[(z-a)^{m\zeta}].\end{aligned} \tag{8.245}$$

One may also choose a desired value for the fractional derivative power $\zeta \in (0,1]$. **NOTE:** it is in the choice of ζ that this approach has its advantage. To see this, consider the results discussed regarding the FLOM, that is, fractional lower order moments, one can then select the value of the fractional derivative ζ to ensure that the largest polynomial power on the truncated series given by $n\zeta$ is selected to produce a finite expectation with respect to the heavy-tailed distribution. This can always be achieved where $\zeta \downarrow 0$ as $n \to \infty$.

2. Take the expectation with respect to the heavy-tailed severity model to obtain the approximation given in the α-stable setting discussed in this section according to

$$\begin{aligned}&\mathbb{E}_{S_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0))}[\widehat{g}(Z)] \\ &= \sum_{m=0}^{n} \frac{1}{\Gamma(m\zeta+1)}(D^{\zeta}_{a|z})^m g(z)|_{z=a}\mathbb{E}_{S_\alpha(z;\widetilde{\beta}_n,\widetilde{\gamma}_n,\widetilde{\delta}_n;S(0))}[(Z-a)^{m\zeta}].\end{aligned} \tag{8.246}$$

Then one may recognize that each summand in the series is now corresponding to an FLOM and results may be adopted to get closed-form expressions for the expectations in each summation for particular heavy-tailed models of interest.

As a note on the coherency of such a risk measure, if one formulates this TTCE result in terms of the function $g(z)$ satisfying the admissibility conditions of the SRM results, then this result will also produce a coherent risk measure as it corresponds to a special approximation of the SRM via a fractional series expansion.

CHAPTER NINE

Recursions for Distributions of LDA Models

This chapter develops and presents a range of recursive numerical and statistical evaluation procedures for assessing the properties and providing alternative statistical estimators for the quantities of interest proposed in previous chapters relating to tail functionals of compound processes. The primary interest will be on the estimation of the tail of the compound process and assessment of the asymptotic approximations developed previously. The basic sections proposed in this chapter will consider recursions that fit into one of the following four categories:

1. Recursions for discretized severity n-fold convolutions;
2. Recursions for continuous severity n-fold convolutions;
3. Recursions for discretized severity compound process LDA models; and
4. Recursions for continuous severity compound process LDA models.

9.1 Introduction

The applications of the recursions considered in this chapter, will be in analysis of the results presented in previous chapters regarding closed-form approximations for the tail of a compound process and the resulting risk measure approximations, which are now studied from a statistical perspective. The reason for this is that although these results are important closed-form solutions, in practice, the implementation of these results can be difficult and challenging. The challenges can arise from a number of areas, either because the models considered do not adequately satisfy all the assumptions utilized to obtain the closed-form expansions, but the expansion is still utilized as an approximation. Alternatively, even if the assumptions are satisfied, the higher order terms discarded asymptotically may still contribute significantly in practice for a given quantile level not in the neighbourhood of infinity. In addition, there is generally a lack of knowledge of the rate of convergence of the estimation accuracy of such

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

risk measure and tail functional approximations making it difficult to quantify the statistical uncertainty associated with the approximations for a finite quantile level. Finally, the impact of parameter estimation and its consequences for the use of such estimates in the closed-form approximations, sensitivity and induced bias is also unknown. Therefore, in this chapter, we demonstrate how to incorporate statistical estimation of the tail quantiles with the closed-form tail approximations to improve estimation and knowledge of the uncertainty in these expansions for a given quantile level.

Throughout the statistics, actuarial and risk literature, there have been numerous papers on estimation procedures and recursions for the evaluation of the distribution of a convolution between independent and identically distributed (i.i.d.) random variables, either for a fixed number of losses or for a compound process (random sum). In general, one will be interested in estimation of the compound process LDA single risk distribution over the entire support via recursive solutions. However, a particular focus in this chapter will be consideration of applications of such recursive solutions to efficient evaluation of the tails of the compound distribution. In such settings, we are interested in numerical procedures to improve the estimation of either the fixed number of losses (n) case, $Z_n = X_1 + \ldots + X_n$ with i.i.d. $X_i \sim F(x)$ where one is interested in approximation of

$$\overline{F}_{Z_n}(x) = \overline{F^{(n)*}}(x) = \mathbb{P}\mathrm{r}[Z_n > x], \tag{9.1}$$

or the random number of losses (N) case, $Z_N = X_1 + \ldots + X_N$, with i.i.d. $X_i \sim F(x)$ independent of N, in which one wants to approximate

$$\overline{F}_{Z_N}(x) = \mathbb{P}\mathrm{r}[Z_N > x] = \sum_{n=1}^{\infty} \mathbb{P}\mathrm{r}[N = n]\overline{F}_{Z_N}(x). \tag{9.2}$$

Different asymptotic expansions and indeed different numerical importance sampling-based procedures will be developed depending on whether the problem posed is of for fixed n or random N.

Remark 9.1 (A Note on Monte Carlo Methods) *We observe that the most popular approach in practice to the simulation of an annual loss distribution in an LDA framework involves what we call a standard Monte Carlo approach. This involves first sampling the number of realizations $N = n$. Then conditional on this number of losses, sampling from the severity distribution n losses and aggregating them. This produces a single draw from the compound process distribution. If this is repeated a large number of times, then M samples can be obtained from the distribution $F_{Z_N}(x)$. The reason why more advanced sampling techniques have been developed as alternatives to this simple standard Monte Carlo approach is because when one estimates the tail of the compound process distribution $\mathbb{P}\mathrm{r}[Z_N > x]$ for a very large quantile then the proportion of the sampled points M that actually result in an aggregated annual loss exceeding the level x reduces significantly. This reduction is also a non-linear relationship between the quantile level x and the number of samples achieving the condition. Therefore, one requires very large computational efforts to obtain samples from the tails of the distribution under such simplistic sampling schemes and this has motivated the development of the approaches discussed in this chapter.*

A foreword on notation contained in this chapter, all distributions and densities will be assumed to be continuous unless otherwise stated, at certain stages in the presentation, it will become important to discretize the severity distributions, at which point it will be made explicit when such representations are utilized to avoid confusion. We note that when a continuous distribution is discretized onto an equispaced grid of points in its support, given for some interval $\Delta \in \mathbb{R}^+$ by $0, \Delta, 2\Delta, \ldots$, this is equivalent to working on the space of integers, $0, 1, 2, \ldots$, the resulting discretized distribution will be denoted by $\{\widehat{f}_n\}_{n\in\mathbb{N}}$, where

$\widehat{f}_n = f(n\Delta)$ and $\widehat{f}_n \geq 0$. Note that in some instances, it may also be required to impose a normalization condition on the discretized distribution values where $\sum_{n=1}^{\infty} \widehat{f}_n = 1$.

In addition, we will adopt the convention from the actuarial literature, see a summary in Sundt & Vernic (2009) in which $\mathscr{P}_1$ will denote the class of all univariate distributions on the integers and $\mathscr{P}_{1l}$ the class of all distributions $f \in \mathscr{P}_1$ satisfying the condition that $f(x) = 0$ for all integers $x < l$ and finally, the notation $\mathscr{P}_{1l}$ which denotes the sub-class of distributions in $\mathscr{P}_{1l}$ with a positive mass at l. Analogously, these definitions will also carry through for functions, that is, densities are denoted by the class $\mathscr{P}$ and functions by the label $\mathscr{F}$, for example, densities only known up to normalization. To clarify this point, we consider $\mathscr{F}_l$ for all integers l to be the class of all functions on the integers and $\mathscr{F}_{1l}$ as the set of functions $f \in \mathscr{F}_1$ which satisfy $f(x) = 0$ for all integers $x < l$.

9.2 Discretization Methods for Severity Distribution

It is common in practice for practitioners to discretely approximate a continuous distribution in order to utilize the Panjer recursion. This could involve, for example, rounding all loss amounts to the nearest multiple of a monetary unit Δ, such as the nearest USD 1000. Consider the approximation of the compound process distribution function via discretization in which we do not impose any requirements on the frequency distribution,

$$F_{Z_n}(x) = \sum_{k=0}^{\infty} p_k F^{(k)*}(x), \tag{9.3}$$

where

$$F^{(k)*}(x) = \int_0^x F^{(k-1)*}(x-\tau) f(\tau) d\tau, \tag{9.4}$$

with

$$F^{(0)*}(x) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \tag{9.5}$$

Now under the discretized representation of the distribution over a grid defined by increments of amount Δ, one can proceeds according to the following defined quantities

$$\widehat{f}(k) = \mathbb{P}\mathrm{r}[X_i = k\Delta], \;\; p_k = \mathbb{P}\mathrm{r}[N = k], \;\; \widehat{f}_{Z_N}(k) = \mathbb{P}\mathrm{r}[Z_N = k\Delta], \tag{9.6}$$

to estimate the terms in this convolution via the recursions for the initialization $\widehat{f}_0 = 0$ for integers $k = 0, 1, 2, \ldots$ according to

$$\begin{aligned} \widehat{f}_{Z_N}(n) &= \sum_{k=1}^{n} p_k \widehat{f}^{(k)*}(n), \;\; n \geq 1, \\ \widehat{f}_{Z_N}(0) &= \mathbb{P}\mathrm{r}[Z_N = 0] = \mathbb{P}\mathrm{r}[N = 0] = p_0, \end{aligned} \tag{9.7}$$

where $\widehat{f}^{(k)*}(n)$ is given by

$$\widehat{f}^{(k)*}(n) = \sum_{i=0}^{n} \widehat{f}^{(k-1)*}(n-i) \widehat{f}(i). \tag{9.8}$$

with $\widehat{f}^{(0)*}(0) = 1$ and $\widehat{f}^{(0)*}(n) = 0$ if $n \geq 1$.

As can be seen earlier, one can obtain a discretized representation of the compound process by discretization of the severity model. This can often lead to numerical efficiencies for computations as will be discussed. In the application of several recursive approaches discussed in this chapter, it is typical in practice to discretize the severity distribution as this can significantly simplify the recursive relationship utilized to evaluate the compound process distribution. Therefore, in this section, we first discuss three different discretizations of the severity distribution that have been proposed in the actuarial literature.

In general, given a continuous severity distribution F with density f, we are interested here in representation of this distribution via a discretized representation over a grid of d points denoted by $\Delta_i \in \{\Delta_1, \Delta_2, \ldots, \Delta_d\}$ with each Δ_i selected in an appropriate manner (depending on the discretization approach) from the support of the distribution F. Then the discretized density for the f is represented exactly (up to normalization) at each grid point according to the set of function values $\widehat{f}(\Delta_i) = f(\Delta_i)$. In general, the spacing of the grid of points is selected to ensure a desired precision is attained in the evaluation of the distribution pointwise and may involve linear or non-linear spacings. In this chapter, we are only considering approaches that utilize linear equispaced grid points $0, \Delta, 2\Delta, \ldots$ which we noted is equivalent to working with distribution on the space of integers $0, 1, 2, \ldots$ One should realize that as illustrated earlier, the results for the integers, are then mapped back to the original units via specification of the quantization unit Δ.

As discussed in Rolin *et al.* (2000, chapter 2) and the paper of Walin & Paris (1998), there are numerous approaches one may adopt to discretize a continuous distribution. In cases in which the distribution is originally discrete, we will maintain a notation in which the distribution at a grid point $x\Delta$ will be denoted by $f(x\Delta)$, whereas in cases in which the distribution is continuous and approximated by a discrete representation at a point $x\Delta$, we will denote such cases by $\widehat{f}(x\Delta)$. The following first two methods will be clearly 'local discretization methods' whilst the third method is a 'global discretization method'.

9.2.1 DISCRETIZATION METHOD 1: ROUNDING

The approach developed in Gerber & Jones (1976) involves a simple rounding mechanism. This approach, whilst simple to apply, has the drawback that it fails to preserve any of the moments of the continuous distribution, which may be undesirable in some applications. The approach is defined according to the following approximation:

$$\begin{aligned} \widehat{f}(0) &= F\left(\frac{\Delta}{2}\right), \\ \widehat{f}(x\Delta) &= F\left(x\Delta + \frac{\Delta}{2}\right) - F\left(x\Delta - \frac{\Delta}{2}\right), \ \forall x \in \{1, 2, \ldots\}. \end{aligned} \tag{9.9}$$

Note that here the value 0 in the arguments is utilized because there is indeed a value of the function at the origin which in the discrete setting will be included in the support.

The other simple approaches to rounding proposed in the literature by Panjer *et al.* (1992) are simply the floor and ceil functions, given in the ceil case by

$$\begin{aligned} \widehat{f}(0) &= 0, \\ \widehat{f}(x\Delta) &= F(x\Delta) - F(x\Delta - \Delta), \ \forall x \in \{1, 2, \ldots\}, \end{aligned} \tag{9.10}$$

and in the floor case by

$$\begin{aligned} \widehat{f}(0) &= F(\Delta), \\ \widehat{f}(x\Delta) &= F(x\Delta + \Delta) - F(x\Delta), \ \forall x \in \{1, 2, \dots\}. \end{aligned} \tag{9.11}$$

Note that the final values of these discretization methods may need to be normalized for some applications.

In Example 9.1, we plot three different examples for the discretization methods on this gamma model.

EXAMPLE 9.1

In this simple example, we consider the distribution function for a gamma severity distribution with $X \sim Gamma(1, 1)$ with a step size of $\Delta = 0.25$ between the interval $[0, 5]$. The result of the application of this discretization is provided in Figure 9.1.

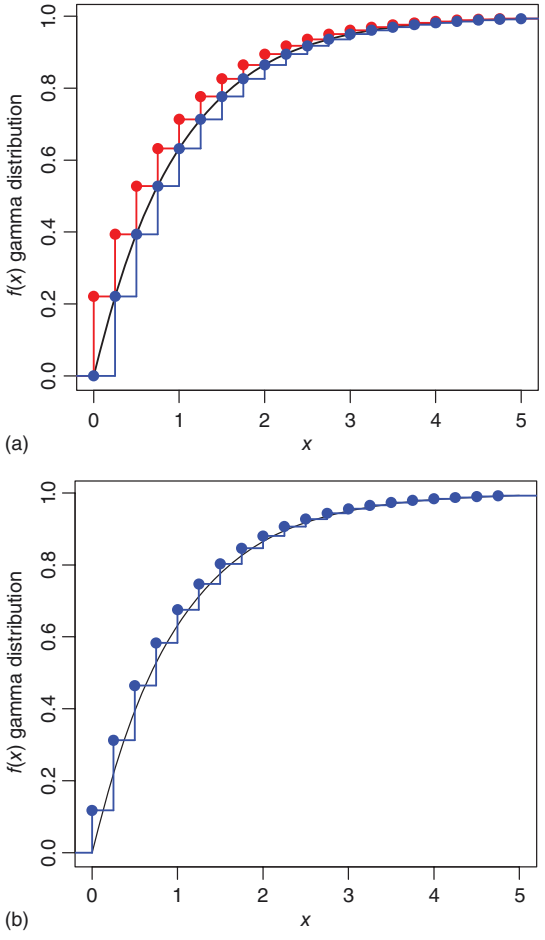

FIGURE 9.1 Plots of the discretized gamma distribution under the following cases: (a) upper and lower rounding rules; (b) midpoint rounding rule.

EXAMPLE 9.2 Rounding Method Weibull Distribution

In this simple example, we consider the distribution function for a Weibull severity distribution with $X \sim Weibull(1, 5)$ with a step size of $\Delta = 0.5$ between the interval $[0, 10]$. The result of the application of this discretization is provided in Figure 9.2.

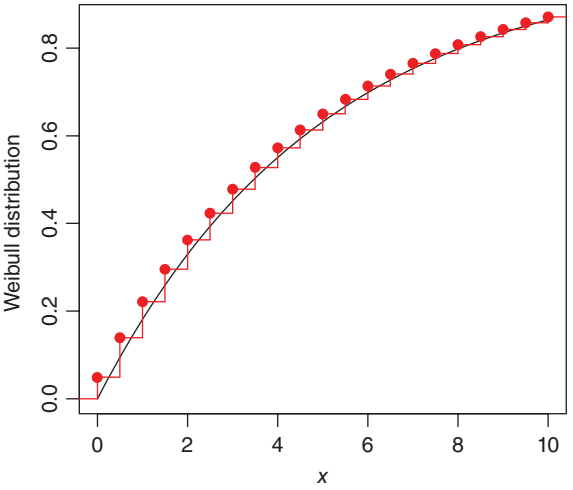

FIGURE 9.2 Discretization midpoint method of rounding for Weibull distribution function severity model.

9.2.2 DISCRETIZATION METHOD 2: LOCALIZED MOMENT MATCHING

In Gerber (1982), an approach was developed which involves a localization of the standard technique of moment matching which produces a system of non-linear equations that can be solved using the Lagrange formula locally for each set of grid points. Consider an interval $(x\Delta, x\Delta + 2\Delta]$ and associate local masses to be solved for at the left grid points denoted by $\{p_0(x\Delta), p_1(x\Delta), p_2(x\Delta)\}$ in which the following system of equations must be solved for the zeroth, first and second local moments:

$$p_0(x\Delta) + p_1(x\Delta) + p_2(x\Delta) = \int_{x\Delta}^{x\Delta+2\Delta} dF(x),$$

$$(x\Delta)p_0(x\Delta) + (x\Delta + \Delta)p_1(x\Delta) + (x\Delta + 2\Delta)p_2(x\Delta) = \int_{x\Delta}^{x\Delta+2\Delta} xdF(x),$$

$$(x\Delta)^2 p_0(x\Delta) + (x\Delta + \Delta)^2 p_1(x\Delta) + (x\Delta + 2\Delta)^2 p_2(x\Delta) = \int_{x\Delta}^{x\Delta+2\Delta} x^2 dF(x).$$

The solution to this system of equations for the masses $\{p_0(x\Delta), p_1(x\Delta), p_2(x\Delta)\}$ is obtained in closed form according to

$$p_j(x\Delta) = \int_{x\Delta}^{x\Delta+2\Delta} \prod_{i \neq j} \frac{\tau - x\Delta - i\Delta}{(j-i)\Delta} dF(\tau), \ \forall j \in \{0, 1, 2\}. \tag{9.12}$$

The system of equations above can be evaluated in closed form in some cases; in other cases, one needs to use, for instance, a quadrature approximation. For instance, one could approximate the integrals via the trapezoidal (trapezium rule) using the result

$$\int_{x_0}^{x_0+x\Delta} f(x)dx \approx f(x_0)\Delta x + \frac{1}{2}f'(x_0)\Delta x^2 + \frac{1}{6}f''(x_0)\Delta x^3 + \ldots$$
$$= \frac{1}{2}(f(x_0) + f(x_0+\Delta x)\Delta x + O(\Delta x^3)), \tag{9.13}$$

with the choices $f(x) = f_X(x)$, $f(x) = xf_X(x)$ or $f(x) = x^2 f_X(x)$ utilized.

Remark 9.2 *Note that we did not explicitly call these probability masses because the solution can actually produce negative results, which are clearly not usable for this discretization method—making this approach sometimes untenable, depending on the grid points and model. To obtain the final discretized distribution, simply add the masses at each grid point to get the discretized distribution approximation. Once the sequence of n discretized values $\{p_j\}_{j=0}^n$ are obtained, they can be turned into values $\{\widehat{f}(j)\}_{j=1}^n$ by making sure they are all positive and normalized. Note that the final values of these discretization methods may need to be enforced to be positive and normalized for some applications.*

EXAMPLE 9.3 Localized Moment-Matching Method Gamma Distribution

As in the rounding method above, we consider the same severity model from a gamma distribution with $X \sim Gamma(1,1)$ with a step size of $\Delta = 0.5$ between the interval $[0,5]$. The result of the application of this local moment matching up to order two produces a discretization provided in Figure 9.3.

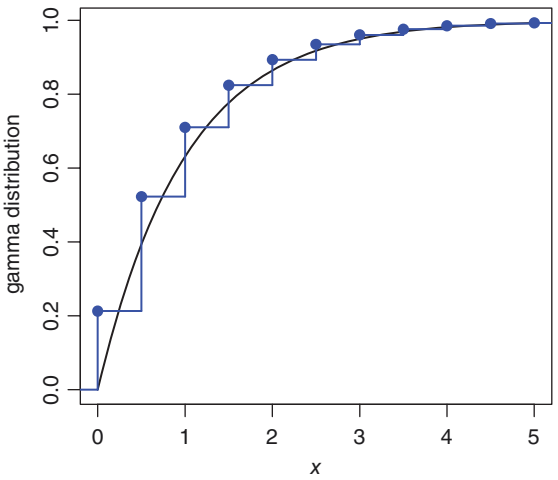

FIGURE 9.3 Discretization method of local moment matching for gamma distribution function severity model.

9.2.3 DISCRETIZATION METHOD 3: LIOYD'S ALGORITHM

This approach was popularized in the signal processing and engineering literature and named after Stuart Lloyd, see Lloyd (1982). This algorithm is a form of Voronoi iteration or relaxation for finding evenly spaced sets of points in subsets of Euclidean spaces and partitions of these subsets into well-shaped and uniformly sized convex cells. In this regard, it is closely related to the concept of k-means clustering. It is well known that in one dimension, Lloyd's algorithm converges to a centroidal Voronoi diagram. This algorithm is typically applied to settings where data is available and one wants to 'group' or 'cluster' the data by partitioning the convex hull of the data. In the setting, we consider here the Lloyd algorithm is applied in a non-standard way, not to data, but instead to the discretization of a known distribution function. In this case, the points are selected as well as the mass at the discrete points to minimize a squared error criterion and, hence, to compute the discretization of the approximate probability density function (pdf) of a severity model denoted by the approximation $\widehat{f}_X(z)$. We can achieve this using the Lloyd algorithm, which minimizes the mean-square error and in the process produces a set of grid points $\{\Delta_i\}_{i=1}^N$ for which we have probability masses $\{\widehat{f}_i\}_{i=1}^N$.

Under this discretization method, the grid points are not required to be uniformly spaced. This will have implications for utilization of this method in calculating integrals which should be considered. The algorithm is detailed by the following steps for N discrete grid points:

Algorithm 9.1 Lloyd's algorithm

1. *Initialization.*

(a) Choose the initial 'quantization' levels, that is, discretization steps $\Delta_i^{(0)}, \quad i = 1, \ldots, N$*;*

(b) Set the quantization boundaries $b_i^{(0)} = \frac{\Delta_{i+1}^{(0)} - \Delta_i^{(0)}}{2}, \quad i = 1, \ldots, N-1$*;*

(c) Construct the initial density function: $\widehat{f}_X(x) = \sum_{i=1}^N \widehat{f}_i \delta_{\Delta_i^{(0)}}(x)$.

(d) Set the initial distortion $D^{(0)} = \infty$.

2. *Update rule: While* $|D^{(k)} - D^{(k-1)}| > \varepsilon$ *perform the following steps:*

(a) Quantizer levels: $\Delta_i^{(k+1)} = \dfrac{\int_{b_{i-1}^{(k)}}^{b_i^{(k)}} x \widehat{f}_X(x)\,dx}{\int_{b_{i-1}^{(k)}}^{b_i^{(k)}} \widehat{f}_X(x)\,dx}$.

(b) Quantizer boundaries: $b_i^{(k+1)} = \frac{\Delta_{i+1}^{(k+1)} - \Delta_i^{(k+1)}}{2}$.

(c) Distortion: $D^{(k+1)} = \sum_{i=1}^N \int_{b_{i-1}^{(k+1)}}^{b_i^{(k+1)}} (x - \Delta_i^{(k+1)})^2 \widehat{f}_X(x)\,dx$.

(d) Increment counter: $k = k + 1$.

Note that the final values of these discretization methods may need to be normalized for some applications.

9.2.4 DISCRETIZATION METHOD 4: MINIMIZING KOLMOGOROV STATISTIC

A simple global approach based on the minimization of the Kolmogorov distance is discussed in Rolin *et al.* (2000, chapter 2) and the error this discretization approach has on a compound process model (LDA single risk) for such a discretization approach has been studied in

Embrechts *et al.* (1993) and Grübel & Hermesmeier (2000). Basically, the approach involves selection of each function value $\widehat{f}(x\Delta)$ at each grid point $x\Delta$, for $x \in \{0, 1, 2, \dots\}$ according to minimization of the Kolmogorov distance between two random variables distribution functions,

$$d_K(X, Y) = \max_x |F_X(x) - F_Y(x)|.$$

In the case considered here, one considers $F_X(x)$ the continuous severity distribution and $F_Y(x) = \widehat{F}_X(x)$ the discrete severity distribution. One can then set the values for the grid by selecting Δ and the masses for $\widehat{F}_X(x)$ are obtained by minimization of this distance measure. The solution to the selection of the discretized distribution that minimizes this distance is then obtained numerically. Note that the final values of these discretization methods may need to be normalized for some applications.

9.3 Classes of Discrete Distributions: Discrete Infinite Divisibility and Discrete Heavy Tails

If one considers a discretized continuous or a discrete distribution $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ which satisfies the constraint that $\widehat{f}_n \geq 0$ for all $n \in \mathbb{N}$ which is normalized with $\sum_n \widehat{f}_n = 1$, then it is first useful to recall the following definition of a non-degenerate discrete distribution.

Definition 9.1 (Non-Degenerate Discrete Distribution) *A discrete distribution that is normalized is called non-degenerate if* $\widehat{f}_n < 1$ *holds for all* $n \in \mathbb{N}$.

In addition, it will be useful to define the class of discrete distributions known as the Panjer class.

Definition 9.2 (General Panjer Class of Discrete Distributions) *A discrete distribution* $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ *with parameters* $a, b \in \mathbb{R}$ *and order* $k \in \mathbb{N}$ *is in the Panjer class if* $\widehat{f}_n = 0$ *for all* $n \leq k - 1$ *and the probabilities satisfy the recursion*

$$\widehat{f}_{n+1} = \left(a + \frac{b}{n+1}\right) \widehat{f}_n, \tag{9.14}$$

for all $n \geq k$. *Such a class of discrete distribution that satisfies this condition is known as the Panjer class family with parameters* a, b *and order* k.

There will be further discussion on these classes of discrete distributions in the following sections. In this section, we will also state a representation of such distributions according to a differential equation in terms of its probability-generating function, see discussion in Hess *et al.* (2001). It will be useful to first define the probability-generating function given in Definition 9.3.

Definition 9.3 (Probability Generating Function) *A random variable* X *with discrete probability mass distribution* $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ *is characterized by the probability-generating function (p.g.f.) given by*

$$m_X(z) := \mathbb{E}[z^X] = \sum_{n=0}^{\infty} \widehat{f}_n z^n, \tag{9.15}$$

where given the p.g.f., denoted by $m_X(z)$, one obtains the probabilities by differentiation and evaluation at the origin according to

$$\widehat{f}_n = \frac{1}{n!}\frac{d^n}{dz^n} m_X(z)\bigg|_{z=0} = \frac{1}{n!} m_X^{(n)}(0), \ \forall n \in \mathbb{N}. \tag{9.16}$$

One can then show that the following representational equivalence for degenerate discrete severity models given in Theorem 9.1.

Theorem 9.1 (Panjer Class k and differential equation for p.g.f.) *If a random variable X has discrete distribution $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ that is non-degenerate, then the following statements are equivalent:*

1. *$\{\widehat{f}_n\}_{n\in\mathbb{N}}$ is in the Panjer a, b class of order k.*
2. *The p.g.f. satisfies the differential equation given by*

$$(1 - az)m_X^{(n+1)}(z) = ((n+1)a + b)m_X^{(n)}(z), \tag{9.17}$$

for $z \in [0, 1)$ and initial condition $m_X^{(j)}(0) = 0$ for all $j \leq k - 1$.

In addition to the Panjer class of distributions, it will sometimes be of relevance to consider lower- and upper-truncated Poisson distributions and in particular convolutions of such random variables when they are considered independent. The class of upper- and lower-truncated Poisson n-fold convolution distributions was studied in Huang & Fung (1993) and is denoted as the family of D-distributions. Several properties of this family of distributions are known such as the expressions for integer moments, see Huang & Fung (1993).

Definition 9.4 (D-Distributions: n-Fold Convolutions of Upper- and Lower-Truncated Poisson) *Consider n independent counting random variables each i.i.d. from an upper-truncated and a lower-truncated Poisson distribution such that*

$$N_i \sim Poisson(\lambda_i)\mathbb{I}[N_i \in \{n_{\min,i}, n_{\max,i}\}], \tag{9.18}$$

for some positive integers $n_{\min}$ and $n_{\max}$ satisfying $0 < n_{\min,i} < n_{\max,i} < \infty$. Then the distribution of the convolution given by

$$N_n = \sum_{k=1}^{n} N_k \sim D\text{-distribution}(n, \mathscr{L}, \mathbf{\Lambda}), \tag{9.19}$$

where the D-distribution is characterized by the discretely supported distribution given by

$$\Pr(X = k) = \prod_{i=1}^{n} e(N_i, M_i; \lambda_i)^{-1} D(x, n; \mathscr{L}\Lambda)\frac{1}{x!}, \tag{9.20}$$

with

$$x \in \left\{ \sum_{i=1}^{n} n_{\min,i}, \sum_{i=1}^{n} n_{\min,i} + 1, \ \ldots \ , \sum_{i=1}^{n} n_{\max,i} \right\}, \tag{9.21}$$

and the D numbers given by

$$D(x, n; \mathscr{L}, \Lambda) = \sum_{\boldsymbol{y} \in \boldsymbol{Y}} \binom{x}{y_1, y_2, \ldots, y_n}, \tag{9.22}$$

with

$$\begin{aligned}
\boldsymbol{Y} &= \left\{ \boldsymbol{y} : \boldsymbol{y} = (y_1, y_2, \ldots, y_n), x = \sum_{i=1}^{n} y_i \right\}, \\
\mathscr{L} &= \{(n_{\min,i}, n_{\max,i}) : n_{\min,i} < n_{\max,i}, n_{\min,i} \in \mathbb{N}^+, n_{\max,i} \in \mathbb{N}^+\}, \\
\Lambda &= \{\lambda_i : \lambda_i > 0\},
\end{aligned}$$

and $e(\cdot)$ the incomplete exponential function with parameter λ given by

$$e(N, M; \lambda) = \begin{cases} \sum_{i=N}^{M} \frac{\lambda^i}{\mathrm{i}!}, & \lambda > 0, 0 \leq N < M,\ N, M \in \mathbb{N}, \\ \sum_{i=0}^{M} \frac{\lambda^i}{\mathrm{i}!}, & \lambda > 0, N = -1, -2, -3, \ldots, \\ 0, & otherwise. \end{cases} \tag{9.23}$$

It will also be useful in sections of this chapter to define the special class of discrete distributions $\{\widehat{f}_n\}_{n \in \mathbb{N}}$ which will have the property that they are infinitely divisible, as characterized in Theorem 9.2, see discussions in Steutel & Van Harn (2003).

Theorem 9.2 (Characterizing the Infinitely Divisible Distributions) *The following properties exist for members of the class of infinitely divisible distributions:*

1. *A distribution concentrated on a Dirac mass is infinitely divisible.*
2. *The class of infinitely divisible distributions is closed under the operation of convolution.*
3. *The class of infinitely divisible distributions is closed under linear translations.*
4. *The class of infinitely divisible distributions is closed under constant scalings.*
5. *A mixed Poisson distribution is infinitely divisible if the mixing distribution is also infinitely divisible.*
6. *A mixed Poisson distribution with infinitely divisible mixing distribution can be expressed as a compound Poisson distribution with severity distribution in the class $\mathscr{P}_{11}$.*
7. *A compound distribution is infinitely divisible if its counting distribution (frequency distribution) is infinitely divisible and in the class of distributions $\mathscr{P}_{10}$;*
8. *An infinitely divisible distribution in the class $\mathscr{P}_{10}$ has a positive probability in zero;*
9. *A non-degenerate distribution f in the class $\mathscr{P}_{10}$ is infinitely divisible if and only if it can be expressed as a compound Poisson distribution with a severity distribution g which is in the class $\mathscr{P}_{11}$.*

Random variables which are distributed according to a discrete distribution which is infinitely divisible have the advantage that when evaluating sums of such random variables under an independence assumption will result in simple expressions for the form of the resulting compound distribution. This is explored in the following sections through the De Pril transforms for such classes of random variable.

To determine if a discrete distribution is infinitely divisible, one can utilize the necessary and sufficient condition for a discrete distribution to be infinitely divisible given in Theorem 9.3, see Katti (1967).

Theorem 9.3 (Discrete Infinitely Divisible Distributions: Necessary and Sufficient Conditions) *Consider a discrete distribution* $\{\widehat{f}_i\}_{i=1}^{N}$ *for* $i = 0, 1, 2, \ldots$ *with* $\widehat{f}_0 \neq 0$ *and* $\widehat{f}_1 \neq 0$. *Then the necessary and sufficient condition for* $\{\widehat{f}_i\}_{i=1}^{N}$ *to be infinitely divisible is that it must satisfy the recursion given below, which must be strictly non-negative for all* $i \in \mathbb{N}$.

$$\pi_i = i\frac{\widehat{f}_i}{\widehat{f}_0} - \sum_{j=1}^{i-1} \pi_{i-j} \frac{\widehat{f}_j}{\widehat{f}_0} \geq 0. \tag{9.24}$$

A corollary of this result discussed in Bondesson *et al.* (1996, corollary 2.6) is given as follows.

Corollary 9.1 *If a discrete distribution* $\widehat{f}_n$ *on positive integers* n *is of the form* $\widehat{f}_n = (n+1)c_n$ *for some sequence* c_n *which is completely monotone, then the distribution is infinitely divisible.*

To validate that a discrete distribution is in the class of infinitely divisible distributions, it was later shown in Warde & Katti (1971) that it is sufficient to ensure the following condition provided in Theorem 9.4 holds.

Theorem 9.4 (Discrete Distribution Infinite Divisibility Sufficient Condition 1) *A discrete distribution* $\{\widehat{f}_n\}_{n\in\mathbb{N}}$, *with* $\widehat{f}_0 \neq 0$, $\widehat{f}_1 \neq 0$ *is infinitely divisible if the ratios* $\frac{\widehat{f}_i}{\widehat{f}_{i-1}}$ *for* $i = 1, 2, \ldots$ *form a monotone increasing sequence.*

A second alternative sufficient condition that can be considered for discrete distributions with support on the non-negative integers, as given in 9.5, see Steutel (1973).

Theorem 9.5 (Discrete Distribution Infinite Divisibility Sufficient Condition 2) *A discrete distribution* $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ *on the non-negative integers, with* $\widehat{f}_0 \neq 0$, *is infinitely divisible if it satisfies the recursion*

$$\widehat{f}_{n+1} = \sum_{k=0}^{n} q_k \widehat{f}_{n-k}, \tag{9.25}$$

for $q_k \geq 0$ *for all* $k = 1, 2, 3, \ldots$

Remark 9.3 *Distributions that satisfy the above are also known to be of compound geometric distribution form.*

Yet a third sufficient condition can be stated for discrete infinitely distributed random variables on the positive integers, given in Theorem 9.6, see discussion in Steutel (1973).

Theorem 9.6 (Discrete Distribution Infinite Divisibility Sufficient Condition 3) *A discrete distribution* $\{\widehat{f}_n\}_{n\in\mathbb{N}}$ *on the non-negative integers with a distribution which is log-convex, is infinitely divisible if it satisfies that*

$$\widehat{f}_{n+1}\widehat{f}_{n-1} \geq \widehat{f}_n^2, \tag{9.26}$$

for all $n = 1, 2, \ldots$

It is also worth noting that it can be shown that the class of discrete distributions satisfying Corollary 9.1 are comprised the class of mixtures of negative binomial distributions of order two.

Having defined the notion of discrete infinite divisibility, it is natural to consider the question, are there any cases in which one can take a distribution which is infinitely divisible with support $(0, \infty)$ and discretize it to a distribution which takes integer support and preserves the infinite divisibility?

Bondesson *et al.* (1996) studied this type of question under a particular form of discretization based on simple rounding by taking the integer component. That is, consider a random variable X decomposed as its integer component $[X]$ and remainder fractional part $\{X\}$ given by $X = [X] + \{X\}$. In this setting, they were able to state the following result in Theorem 9.7. Before stating this result, we will briefly recall the definition of a log-concave density function for an operational risk (OpRisk) severity model, see Definition 9.5.

Definition 9.5 (Log-Concave Density Functions) *A continuous loss random variable X has a density $f_X(x)$ which is said to be log-concave if it can be expressed as follows:*

$$f_X(x) = \exp(\phi(x)), \tag{9.27}$$

where $\phi(x)$ is a concave function.

Densities that are log-concave satisfy the following conditions:

1. $\ln f_X(\lambda x + (1-\lambda)y) \geq \lambda \ln f_X(x) + (1-\lambda) f_X(y)$;
2. Analogously, $f_X(\lambda x + (1-\lambda)y) \geq f_X(x)^{\lambda} f_X(y)^{(1-\lambda)}$;
3. $f_X\left(\frac{1}{2}(x+y)\right) \geq \sqrt{f_X(x) f_X(y)}$;
4. In Ibragimov (1956), it was shown that a density on $\mathbb{R}$ will be log-concave if and only if when it is convolved with a unimodal density, the resulting convolved density is again unimodal.

Theorem 9.7 (Continuous Infinite Divisibility to Discrete Infinite Divisibility) *If a loss random variable X with support on $(0, \infty)$ has a log-concave density then both X and the integer supported random variable given by $[X]$ are characterized by infinitely divisible distributions.*

In addition, it was shown in Bondesson *et al.* (1996, theorem 3.3) that one can characterize the classes of infinitely divisible severity distributions for which the integer rounded discrete distributions will also be infinitely divisible as those satisfying the result in Theorem 9.8.

Theorem 9.8 (Characterizing Infinitely Divisible Distributions: Continuous and Discrete) *Consider a continuous loss random variable with a severity density that satisfies the representation given by*

$$f_X(x) = (x + a)h(x), \tag{9.28}$$

where $a \geq 0$ and $h(x)$ is a completely monotone function. In this case, the severity distribution is infinitely divisible. In addition, $[X]$ the discretized loss random variable on the integers is infinitely divisible for such a model when $a \geq 1$.

In mentioning the class of discrete infinitely divisible distributions, one can then ask what types of models may be suitable to consider in this class for OpRisk modeling of severity.

Before proceeding to discuss such models of relevance to OpRisk settings, we first recall the definition of the discrete Sibuya distribution given in Definition 9.6, see discussion in Devroye (1993).

Definition 9.6 (Discrete Sibuya Distribution) *A random variable S is said to be distributed according to the Sibuya distribution with parameter γ if it has a p.g.f. given by*

$$m_S(z) = 1 - (1-z)^{\gamma}, \tag{9.29}$$

for $\gamma \in (0, 1]$. In addition, it can, therefore, be shown that the discrete probability mass function is given by

$$\mathbb{P}\mathrm{r}(S = n) = \begin{cases} \dfrac{\gamma(1-\gamma)\cdots(n-1-\gamma)}{n!}, & n > 1, \\ \gamma, & n = 1. \end{cases} \tag{9.30}$$

In turns out that there is a class of discrete severity distributions that form the analogue of the α-Stable severity model, for the discrete support case as given in Definition 9.7, see discussion in Steutel & Van Harn (2003) and Christoph & Schreiber (1998a).

It is useful to recall that α-stable random variables, which are discussed in Chapter 3, satisfy the condition that for strictly stable random variables $X_i \in \mathbb{R}$ with $i = 0, 1, 2, \ldots, n$, one has the characterization given by

$$X_1 + \ldots + X_n \stackrel{d}{=} c_n X_0 + d_n, \tag{9.31}$$

for some sequence $c_n = n^{\frac{1}{\alpha}}$ and (strict stability involves $d_n = 0$). The tail index coefficient $\alpha \in [0, 2]$ is the index of stability, dictating how heavy the tails of the distribution will be, with light tails if α is close to 2 and heavy tailed as α decreases towards 0. In addition, it is known that α must satisfy the condition

$$X \stackrel{d}{=} kX_1 + (1 - k^{\alpha})^{\frac{1}{\alpha}} X_2, \tag{9.32}$$

for $k \in (0, 1)$. It was shown in Christoph & Schreiber (1998a) that one can utilize this condition to re-express the discrete version of the α-stable family of severity models on the integer support which is non-negative.

Definition 9.7 (Discrete Integer Values Stable Family of Severity Models: Class 1) *A discrete non-negative lattice loss random variable X is stable distributed with stability index $\gamma \in (0, 1]$, the discrete analogue of tail index α, if the p.g.f. is represented by*

$$m_X(z) = \exp(-\lambda(1-z)^{\gamma}), \quad |z| \leq 1, \tag{9.33}$$

for some parameter $\lambda > 0$. The probability mass function of a discrete stable loss random variable is given by

$$\mathbb{P}\mathrm{r}(X = k) = (-1)^k e^{-\lambda} \sum_{m=0}^{k} \sum_{j=0}^{m} \frac{m!(\gamma j)!}{(m-j)!j!(\gamma j - k)!k!} (-1)^j \frac{\lambda^m}{m!}, \quad k = 0, 1, 2, \ldots \tag{9.34}$$

It will also be useful to observe the following recursive evaluation available to probabilities for such a distribution given in Lemma 9.1, see Christoph & Schreiber (1998a, theorem 2).

Lemma 9.1 (Recursions for Discrete Stable Probabilities) *If X is a discrete stable severity loss random variable with discrete stable model parameters $\gamma \in (0, 1]$ and λ, then the following recursion holds for the probability evaluations*

$$(k+1)\mathbb{P}\mathrm{r}(X = k+1) = \lambda \sum_{m=0}^{k} \mathbb{P}\mathrm{r}(X = k-m) \frac{(m+1)(-1)^m \gamma!}{(\gamma-(m+1))!(m+1)!}, \tag{9.35}$$

for all $k \in \mathbb{N}^+$ and with $\mathbb{P}\mathrm{r}(X = 0) = \exp(-\lambda)$.

One can also make the following remarks regarding the discrete stable distribution.

Remark 9.4 *The following properties of the discrete stable severity distribution are known:*

- *One can consider the standard characterization of the α-stable distribution given by Equation 9.32 being modified to obtain the discrete stable analogue by replacing kX by the term $k \circ X$ defined by the sum of i.i.d. random variables Z_j satisfying the equality in distribution given by*

$$k \circ X \stackrel{d}{=} \sum_{j=0}^{X} Z_j, \tag{9.36}$$

  for $k \in (0, 1)$ and $\mathbb{P}\mathrm{r}(Z_j = 1) = 1 - \mathbb{P}\mathrm{r}(Z_j = 0) = q$, that is, it admits a mixed Poisson form, where the mixing distribution of the Poisson is known in the literature as the Sibuya distribution.
- *If the discrete stable random variable has a stability index parameter $\gamma = 1$, then it will correspond to the Poisson distribution with intensity parameter λ.*
- *If the discrete stable random variable has a stability index parameter $\gamma < 1$, then the resulting random variable has a distribution which is infinitely divisible, discrete self-decomposable, unimodal and normally attracted to a stable law.*
- *As the discrete stable distribution is in the domain of attraction of a strictly stable random variable, then it also has the following fractional lower order moment properties*

$$\mathbb{E}[X^r] < \infty, \tag{9.37}$$

  for $0 \leq r < \gamma < 1$.
- *The simulation of discrete stable random variables are studied in Devroye (1993). In this paper, it is shown that because the discrete stable distribution can be written as a compound Poisson distribution, then random variable realizations can be drawn from the model*

$$X = \sum_{i=1}^{Y} Z_i, \tag{9.38}$$

  with $Y \sim Poisson(\lambda)$ and i.i.d. discrete random variables Z_i given by the Sibuya distribution with p.g.f. given by

$$m_Z(z) = 1 - (1-z)^{\gamma}. \tag{9.39}$$

- *Hence, one has the discrete stable-distributed random variable* $X \sim S_\gamma(\lambda)$ *if it can be represented in law according to a mixed Poisson distribution*

$$X \sim Poisson\left(\lambda^{\frac{1}{\gamma}} S\right), \tag{9.40}$$

with random variable

$$S \sim Sibuya(\gamma, 1). \tag{9.41}$$

A related family of discrete stable distributions are known as the discrete Linnik distributions first studied in Christoph & Schreiber (1998b), see Definition 9.8.

Definition 9.8 (Discrete Stable Laws: Class 2—Linnik Family) *A discrete severity random variable* X *has a Linnik Law if it has a p.g.f. given by*

$$m_X(z) = \frac{1}{(1 + (1 - z)^\gamma)^\beta}, \tag{9.42}$$

with $\beta > 0$ *and* $\gamma \in (0, 1]$.

Remark 9.5 *It can be shown that the discrete Linnik-distributed severity random variable will also admit a mixed Poisson distribution representation given by consideration a gamma*

$$X \sim Poisson\left(G^{\frac{1}{\gamma}} S\right), \tag{9.43}$$

with random variable

$$S \sim Sibuya(\gamma, 1), \tag{9.44}$$

and independent gamma random variable

$$G \sim Gamma(\beta, 1). \tag{9.45}$$

One can also demonstrate the following useful asymptotic tail behaviours of the discrete stable distribution (class 1), given in Theorem 9.9, see Christoph & Schreiber (1998a, theorem 3).

Theorem 9.9 (Tail Asymptotic of Discrete Stable Distributions) *Consider a discrete stable severity loss random variable* X *with parameters* $\gamma \in (0, 1)$ *and* λ. *Then the following asymptotics hold as* $n \to \infty$

$$\Pr(X = n) = \frac{1}{\pi} \sum_{j=1}^{m} \frac{(-1)^{j+1}}{j!} \lambda^j \sin(\gamma j \pi) B(\gamma j + 1, n - \gamma j) + O(n^{-\gamma(m+1)-1}),\ m < n,$$

with beta function $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. *In addition, one can show that the right tail probability is given by*

$$\Pr(X \geq n) = \frac{1}{\pi} \sum_{j=1}^{m} \frac{(-1)^{j+1}}{j!} \lambda^j \sin(\gamma j \pi) B(\gamma j, n - \gamma j) + O(n^{-\gamma(m+1)}),\quad m < n.$$

9.4 Discretization Errors and Extrapolation Methods

If one chooses to adopt a discretization method for the severity distribution that will then be utilized in recursions for evaluation of convolutions for fixed numbers of losses or in evaluation of recursions for compound process random sums, it is interesting to observe that one can often reduce the discretization error introduced. In particular, the Richardson extrapolation, also known as extrapolation to the limit or acceleration of convergence, can often be performed to reduce discretization errors. This can be particularly relevant when evaluating risk measures based on these discretized annual loss distributions. The approaches discussed in the two articles by Grübel & Hermesmeier (1999, 2000) consider how to increase the accuracy for a given discretization budget through methods that reduce aliasing error and incorporate extrapolation procedures. In particular, the method of the Richardson extrapolation as developed in Richardson (1911) is discussed. The Richardson extrapolation is considered as a sequence acceleration procedure which can improve the rate of convergence of sequences, such as those that may arise in recursive evaluation of convolution distributions on a discretization grid. Examples of the widespread use of the Richardson extrapolation include the method of the Romberg integration Romberg (1955) which combines the Richardson extrapolation to speed up a trapezoidal integration rule for definite integrals.

We first briefly recall how the Richardson extrapolation works and then consider how it may be utilized to improve the accuracy of recursions for compound process recursions. The generic specification of the Richardson extrapolation if provided in Definition 9.9.

Definition 9.9 (Richardson Extrapolation) *Consider approximating a generic density f, for instance, from a convolution, at a particular point x which cannot be evaluated directly. However, one has an approximation that is a function of the discretization effort denoted generically by h which produces an approximation $\widehat{f}_h(x)$ for any $h > 0$ and relates to the true density $f(x)$ according to*

$$\widehat{f}_h(x) = f(x) + Ch^{\alpha} + o(h^{\beta}). \tag{9.46}$$

The Richardson extrapolation then takes the approximations for different h values and combines them in such a manner as to improve the rate of convergence and, therefore, the accuracy of the combined approximation. Assuming that rate α is known but the constant C is intractable, then one can define a new approximation with improved convergence rate of $\widehat{f}_h \to f(x)$ by the following combined approximation known as the Richardson extrapolation of $\widehat{f}_h(x)$ to give

$$\widetilde{f}(x; h, k) := \frac{k^{\alpha}\widehat{f}_h(x) - \widehat{f}_{kh}(x)}{k^{\alpha} - 1}, \tag{9.47}$$

for some h and k, where clever choices of h and k can significantly accelerate the convergence of the new approximation.

The first proposal to utilize extrapolation methods to accelerate insurance-based recursions and evaluations was in the definite integration of a function with respect to a compound process, such as would be required under a spectral risk measure (SRM), see discussion in Embrechts *et al.* (1993). Then it was developed further in the context of compound processes and convolution recursions in the work of Grübel & Hermesmeier (1999) which we discuss briefly later.

The application of the Richardson extrapolation as applicable to the recursions in this chapter will in general follow a four-stage procedure, which is based on a given selected

k, h and selected α for the given approximation method according to Equation 9.50 , as follows:

Algorithm 9.2 Richardson's Extrapolation Algorithm

1. *Discretize the severity distribution input to obtain both $\widehat{f}_{X,h}(x)$ and $\widehat{f}_{X,kh}(x)$ where the h will correspond to the discretization unit Δ as discussed in the methods previously mentioned.*
2. *Given the discretized severity distributions, these can then be considered as inputs to a generic non-linear mapping $\Phi(\cdot)$. The mapping corresponds in this chapter to one of the recursions to be presented throughout this chapter that allows one to evaluate, for instance, the density, distribution or tail of the n-fold convolutions of severity distributions or the evaluation of these quantities for the compound process of an annual loss LDA model given by*

$$\widehat{f}_{Z,h}(x) = \Phi(\widehat{f}_{X,h}(x)), \tag{9.48}$$

   for some mapping $\Phi(\cdot)$ that represents the class of numerical recursion utilized for the evaluation of the intractable density or distribution.
3. *Evaluate the Richardson extrapolation based on approximations $\widehat{f}_{Z,h}(x)$ and $\widehat{f}_{Z,kh}(x)$ given by Equation 9.47 according to*

$$\widetilde{f}_Z(z) := \frac{k^\alpha \widehat{f}_{Z,h}(z) - \widehat{f}_{Z,kh}(z)}{k^\alpha - 1}. \tag{9.49}$$

4. *Evaluate required functionals based on the Richardson extrapolated result for density $\widetilde{f}_Z(z)$ and distribution $\widetilde{F}_Z(z)$ based on the mapped discretized input severity distributions $\widehat{f}_{Z,h}(z)$ and $\widehat{f}_{Z,kh}(z)$, to obtain functional approximations such as*

$$\begin{aligned}
\overline{F}_Z(x) &\approx \int_0^x d\widetilde{F}_Z(x) = \int_0^x \widetilde{f}_Z(x)dx, \\
\overline{F}_X^{(n)*}(x) &\approx \int_0^x d\widetilde{F}_X^{(n)*}(x) = \int_0^x \widetilde{f}_X^{(n)*}(x)dx, \\
\mathrm{SRM}_{x,h}[Z] &\approx \int \psi(x)d\widetilde{F}_Z(x).
\end{aligned}$$

For example, consider the case of the SRM that will be evaluated as the integral of the risk aversion function $\psi(x)$ with respect to the annual loss distribution that was discussed in Cruz *et al.* (2015, chapter 6). Then consider the generic representation

$$\widehat{f}_h(x) = f(x) + Ch^\alpha + o(h^\beta). \tag{9.50}$$

In this context, the extrapolation to the limit relation of this type for the approximation of the SRM would be given by considering an approximation integral expansion of the form

$$\begin{aligned}
\mathrm{SRM}_{\alpha,h}[Z] &:= \int \psi(x)dF_{Z,h}(x) \\
&= \int \psi(x)d\Phi(F_{X,h}(x)) \\
&= \int \phi(x)d\Phi(F_X(x)) + \zeta(\Phi; F_X, \psi)h^\alpha + o(h^\beta),
\end{aligned} \tag{9.51}$$

as $h \downarrow 0$ and $\beta \geq \alpha > 0$ and function $\zeta(\Phi; F_X, \psi) \neq 0$. Examples are provided for particular instances of α and β rates in Embrechts *et al.* (1993) in the case of integration results.

In the case that one is primarily interested in just the discretization density say from an n-fold convolution of the severity $\widehat{f}_{X,h}^{(n)*}$, then the key results are developed for the Richardson extrapolation methods in Grübel & Hermesmeier (1999). In particular, if one considers that a density exists for $\widehat{f}_X^{(n)*}$ for the measure $\nu_X^{(n)*}$, then one way of considering the convergence of $\widehat{f}_{X,h}^{(n)*}$ to $f_X^{(n)*}$ as discretization error is diminished ($h \downarrow 0$) is to consider the approximation

$$\widehat{f}_{X,h}(kh) = \frac{1}{h}\nu_{X,h}(\{kh\}) + g(kh)h^{\alpha} + O(h^{\beta}), \tag{9.52}$$

which holds uniformly in integer discretization steps $k \in \mathbb{J}^+$ as $h \downarrow 0$ for some function $g(\cdot)$ which depends on the mapping $\phi(\cdot)$ and measure input. This could be combined into the Richardson extrapolation procedure as long as $\beta > \alpha$. Details of particular examples of function g and rates α and β are provided in detail in Grübel & Hermesmeier (1999, section 3).

9.5 Recursions for Convolutions (Partial Sums) with Discretized Severity Distributions (Fixed *n*)

In the following sections, we will discuss and explore recursions for partial sums and compound processes. The three main classes of recursion considered are known as the Panjer, De Pril and the method of Kornya's approximation, see discussions and comparisons in Kuon *et al.* (1987).

Here we consider the recursive evaluation of an n-fold convolution $F_{Z_n}(x)$ in the case in which the severity distribution has been discretized onto an equispaced grid (according to one of the approaches presented previously), w.l.o.g. over integer values $x = 1, 2, 3 \ldots$ In this case, it is proven in Sundt & Vernic (2009, theorem 2.8) that a recursion for the evaluation of the discretized density arising from the n-fold convolution distribution $\widehat{F}_{Z_n}(x)$ is achieved according to the result in Theorem 9.10.

Consider a discrete density $\widehat{f} \in \mathscr{P}_{10}$, then the evaluation of the n-fold convolution of such a distribution $\widehat{f}_{Z_n}(x) = \widehat{f}^{(n)*}(x)$ is obtained according to the recursive relationship developed in Theorem 9.10.

Theorem 9.10 (Discretized Severity Distribution n-Fold Convolution Recursions) *The n-fold convolution for the density $\widehat{f}_{Z_n}(x) = \widehat{f}^{(n)*}(x)$, with severity distribution satisfying the condition that $\widehat{f}(x) = 0,\ \forall x < 0$, can be evaluated according to the following recursion*

$$\widehat{f}_{Z_n}(x) = \frac{1}{\widehat{f}(0)} \sum_{y=1}^{x} \left((n+1)\frac{y}{x} - 1 \right) \widehat{f}(y)\widehat{f}_{Z_n}(x-y), \quad \forall x \geq 1, \tag{9.53}$$

with initialization given by

$$\widehat{f}_{Z_n}(0) = \widehat{f}(0)^n. \tag{9.54}$$

To understand the initialization stage in this recursion, consider the two-fold convolution between a distribution $\widehat{f} \in \mathscr{P}_{10}$ and itself which is then given by

$$(\widehat{f} * \widehat{f})(z) = \sum_{x=-\infty}^{\infty} \widehat{f}(x)\widehat{f}(z-x), \ \forall z = 0, 1, 2, \ldots \tag{9.55}$$

As one has that $\widehat{f}(x) = 0$ for $x < 0$ and $\widehat{f}(z-x) = 0$ for $x > z$ because of the membership of this density in the class $\mathscr{P}_{10}$, then one gets the sum

$$(\widehat{f} * \widehat{f})(z) = \sum_{x=0}^{\infty} \widehat{f}(x)\widehat{f}(z-x), \ \forall z = 0, 1, 2, \ldots, \tag{9.56}$$

and in particular, at the origin, one obtains

$$(\widehat{f} * \widehat{f})(0) = \widehat{f}(0)^2. \tag{9.57}$$

When this is extrapolated to the n-fold convolution, one obtains the condition, $\widehat{f}_{Z_n}(0) = \widehat{f}(0)^n$. Next, to understand where the recursive relationship is derived from, we adopt the approach in Sundt & Vernic (2009), where one considers the addition of an auxiliary random variable Y with the density $\widehat{f}$ which is independent of the random variable Z_n which has density $\widehat{f}^{(n)*}$. Then one can show via an argument of symmetry that the following expression holds for all $z_n \in \{1, 2, \ldots\}$ according to

$$\sum_{y=0}^{z_n} \left((n+1)\frac{y}{z_n} - 1 \right) \widehat{f}(y)\widehat{f}^{(n)*}(z_n - y) = 0, \tag{9.58}$$

from which one can obtain a recursive expression for the solution for $\widehat{f}_{Z_n}(x) = \widehat{f}^{(n)*}(x)$ as detailed in Theorem 9.10.

If there is a known upper bound on the severity distribution because of the application of an insurance policy or the total liability that one may be exposed to for a given type of OpRisk as discussed in Cruz *et al.* (2015, chapter 16), then in this case, one can develop a recursion for the truncated discretized severity distribution given in Theorem 9.11, see details in Sundt & Vernic (2009, theorem 2.9).

Theorem 9.11 (Discretized Truncated Severity n-Fold Convolution Recursions) *The n-fold convolution for the density $\widehat{f}_{Z_n}(x) = \widehat{f}^{(n)*}(x)$, with severity distribution $\widehat{f} \in \mathscr{P}_1$, and furthermore, there exists a finite k such that $k = \max\{x : \widehat{f}(x) > 0\} < \infty$, then one can evaluate the convolved distribution according to the following recursion*

$$\widehat{f}_{Z_n}(x) = \frac{1}{\widehat{f}(k)} \sum_{y=1}^{nk-x} \left(\frac{y(n+1)}{nk-x} - 1 \right) \widehat{f}(k-y)\widehat{f}_{Z_n}(x+y), \ \forall x \in \{nk-1, nk-2, \ldots, 0\},$$

with initialization given at the maximum index k by

$$\widehat{f}_{Z_n}(k) = \widehat{f}(k)^n.$$

In the cases in which there exist a need to evaluate recursively, a sequence of convolutions corresponding to $\{\widehat{f}^{(j)*}\}_{1\leq j\leq n}$ for each $j \in \{1, 2 \ldots, n\}$ for a discrete density $\widehat{f} \in \mathscr{P}_{10}$. In this case, the above convolution identities may be excessively computational compared to a simple recursive evaluation in $j = 1, 2, 3 \ldots, n$ for all $x \in \{0, 1, 2, \ldots\}$ involving

$$\begin{aligned}\widehat{f}^{(j)*}(x) &= \widehat{f} * \widehat{f}^{(j-1)*}(x) \\ &= \begin{cases} \sum_{y=0}^{x} \widehat{f}(y)\widehat{f}^{(j-1)*}(x-y), & \forall j \in \{1,3,5,7,\ldots\}, \\ 2\sum_{y=0}^{(x-1)/2} \widehat{f}^{(j/2)*}(y)\widehat{f}^{(j/2)*}(x-y) & \\ \quad +(\widehat{f}^{(j/2)*}(x/2))^2 \mathbb{I}_{x \text{ even}}, & \forall j \in \{2,4,6,8,\ldots\}. \end{cases}\end{aligned}$$

However, one can improve on the efficiency of evaluation for a sequence of distributions $\{\widehat{f}^{(j)*}\}_{1\leq j\leq n}$ if the additional information is known about the distributions in the convolution. For instance, consider the result in Proposition 9.1, based on Sundt & Vernic (2009, theorem 2.7).

Proposition 9.1 *Consider a density $\widehat{f} \in \mathscr{P}_{10}$ that satisfies, for some a and b, the recursive relationship*

$$\widehat{f}(x) = \left(a + \frac{b}{n}\right) \widehat{f}(x-1), \ \forall x \in \{1, 2, \ldots\}. \tag{9.59}$$

Then the convolution of $\widehat{f}$ with itself $\widehat{f}^{(2)*}$ also satisfies this recursion in Equation 9.59 with a unchanged and the new $\widetilde{b}$ given by $\widetilde{b} = a + 2b$. As a consequence when extrapolated to the case of evaluation of each distribution in the sequence $\{\widehat{f}^{(j)*}\}_{1\leq j\leq n}$, one gets the relationship

$$\widehat{f}^{(j)*}(x) = \left(a + \frac{(a+b)j - a}{x}\right) \widehat{f}^{(j)*}(x-1), \ \forall j, x \in \{1, 2, \ldots\}. \tag{9.60}$$

Having defined these basic n-fold convolution identities for discretized severity models and the sequential version for sequences of increasing n-fold convolutions, we next present a well-known family of transforms that further improve the efficiency of evaluation of convolutions between discretized severity distributions.

9.5.1 DE PRIL TRANSFORMS FOR *n*-FOLD CONVOLUTIONS (PARTIAL SUMS) WITH DISCRETIZED SEVERITY DISTRIBUTIONS

In this section, we introduce the notion of a De Pril transform for a distribution which can then be utilized to devise an efficient recursive relationship for the evaluation of an n-fold convolution. De Pril (1986) and Karl-Heinz (1994) developed a range of recursive identities that were introduced relating to convolutions for partial sums, after discretization of the severity model. It was then later recognized that these identities were highly efficient methods for evaluation of the distribution of an n-fold convolution comprised distributions defined over the non-negative integers with a positive probability at the origin, and in Sundt (2005), Sundt (1998) and Dhaene & Vandebroek (1995), one of the most important of the recursions identified in De Pril's early work was named in his honour as the De Pril transform.

In Definition 9.10, we provide the formal representation of the De Pril transform of a discrete probability density on the non-negative integers with a positive probability mass in zero, $\widehat{f} \in \mathscr{P}_{10}$.

Definition 9.10 (De Pril Transform of a Density) *The De Pril transform, denoted by φ_f, of a discrete probability density on the non-negative integers with a positive probability mass in zero, $\widehat{f} \in \mathscr{P}_{10}$, is given by*

$$\varphi_f(x) = \frac{1}{\widehat{f}(0)} \left[x\widehat{f}(x) - \sum_{y=1}^{x-1} \varphi_f(y)\widehat{f}(x-y) \right], \ \forall x \in \{0, 1, 2, \dots\}, \tag{9.61}$$

with $\varphi_f(0) = 0$.

Remark 9.6 *It can be observed that solving this recursion with respect to $\widehat{f}(x)$ produces*

$$\widehat{f}(x) = \frac{1}{x} \sum_{y=1}^{x} \varphi_f(y)\widehat{f}(x-y), \ \forall x \in \{1, 2, \dots\}, \tag{9.62}$$

and conversely, one can obtain the expression for the De Pril transform in Equation 9.61 by solving Equation 9.62 with respect to $\varphi_f(x)$. Hence, given φ_f and the initial value $\widehat{f}(0)$, then the distribution of $\widehat{f}$ can be evaluated recursively. Therefore, one can show using the property of normalization of the probability density (mass) $\widehat{f}$,

$$\sum_{x=0}^{\infty} \widehat{f}(x) = 1, \tag{9.63}$$

that the De Pril transform φ_f is a unique representation of the distribution $\widehat{f}$.

It is also the case that a distribution in the class $\mathscr{P}_{10}$ will be infinitely divisible if and only if the De Pril transform of its density is non-negative.

9.5.2 DE PRIL'S FIRST METHOD

In this section, we consider the partial sum given by

$$Z_n = \sum_{i=1}^{n} X_j, \tag{9.64}$$

with each X_i independent with distributions $X_i \sim F_i(x)$ and density $f_j(x)$. After discretization, one has probability mass functions for each random variable given by $\{\widehat{f}_i\}_{i=1}^n$. Typically, in OpRisk, one would consider the case that all X_i loss random variables were i.i.d. If each discretized probability mass function $\widehat{f}_j \in \mathscr{P}_{10}$, then one can find the n-fold convolution

$$\widehat{f}_{Z_n}(x) = *_{j=0}^{n} \widehat{f}_j = (\widehat{f}_1 * \widehat{f}_2 * \cdots * \widehat{f}_n)(x), \tag{9.65}$$

using the De Pril transform as displayed below. That is, the real benefit of utilizing the De Pril transform in the context of efficiently evaluating the n-fold convolution of a set of n different discretized severity distributions $\widehat{f}_j \in \mathscr{P}_{10}$ for all $j \in \{1, 2, \dots, n\}$, given by $\widehat{f}^{(n)*}(x) = *_{j=0}^{n} \widehat{f}_j$, is presented by the result in Theorem 9.12, see De Pril (1989). The approach described later for the evaluation of the n-fold convolution is known colloquially as De Pril's first method which involves the steps:

Step 1. For each distribution $\widehat{f}_j$, evaluate the De Pril transform $\varphi_{f_j}(x)$ in Equation 9.62. This step can be simplified by noting that the De Pril transform of the n-fold convolution of a distribution in $\mathscr{P}_{10}$ is n times the De Pril transform of that distribution, see Sundt & Vernic (2009, corollary 6.2).

Step 2. Find the De Pril transform $\varphi_{f^{(n)*}(x)}$ of the convolved distribution $\widehat{f}^{(n)*}(x) = *_{j=0}^{n}\widehat{f}_j$ by simply summing the n De Pril transforms.

Step 3. Find the evaluation of the n-fold convolved distribution $\widehat{f}^{(n)*}(x)$ by using the recursion in Equation 9.62.

Theorem 9.12 (De Pril Transform of an n-Fold Convolution) *The De Pril transform of the convolution of a finite number of discrete densities $\widehat{f}_j \in \mathscr{P}_{10}$ for $j \in \{1, 2, \ldots, n\}$ is given by $\widehat{f}^{(n)*}(x) = *_{j=1}^{n}\widehat{f}_j$ and can be evaluated exactly as the sum of the De Pril transforms of these discrete densities, where*

$$\varphi_{f^{(n)*}}(x) = \sum_{j=1}^{n}\varphi_{f_j}(x) = \frac{1}{x}\sum_{j=1}^{n}\sum_{y=1}^{x}\varphi_{f_j}(y)\widehat{f}_j(x-y),\ \forall x \in \{1, 2, \ldots\}, \tag{9.66}$$

with $\varphi_{f_j}(0) = 0$ for all $j \in \{1, 2, \ldots, n\}$.

As a result of this theorem, one observes that the evaluation of the n-fold convolution can be performed exactly through a linear combination of the De Pril transforms.

Remark 9.7 *Hence, it is clear that the evaluation of an n-fold convolution of discrete densities in $\mathscr{P}_{10}$ can be performed by first obtaining the De Pril transform of each density and then finding the De Pril transform of the convolution through summation. Then given the De Pril transform of the convolution, one trivially obtains the evaluation of the n-fold density through*

$$\widehat{f}^{(n)*}(x) = \frac{1}{x}\sum_{y=1}^{x}\varphi_{f^{(n)*}}(y)\widehat{f}^{(n)*}(x-y),\ \forall x \in \{1, 2, \ldots\}, \tag{9.67}$$

with initial value given by $\widehat{f}^{(n)*}(0) = \prod_{j=1}^{n}\widehat{f}_j(0)$.

9.5.3 DE PRIL'S SECOND METHOD

Consider the case of evaluation of the n-fold convolution given by a set of severity distributions $\widehat{f}_j \in \mathscr{P}_{10}$ for all $j \in \{1, 2, \ldots, n\}$ according to $\widehat{f}^{(n)*}(x) = *_{j=0}^{n}\widehat{f}_j$. Under the second method of De Pril, the idea is to bypass the utilization of the recursive evaluation for each distribution given by applying the recursion in Equation 9.61 and instead to utilize a closed-form expression for evaluation of the De Pril transform of each density given by $\varphi_{f_j}(x)$.

As described in Sundt & Vernic (2009), this involves the following steps for the De Pril second method:

Step 1. For each distribution $\widehat{f}_j$, evaluate the De Pril transform $\varphi_{f_j}(x)$ in closed form, *avoiding the recursive evaluation.*
Find the De Pril transform in closed form by representing each distribution $\widehat{f}_j$ in terms of a compound Bernoulli representation. The compound Bernoulli for discretized distribution $\widehat{f}_j$ will have a frequency π_j and severity $\widehat{f}_j$ component.

- A frequency distribution given by a Bernoulli distribution with probability of success $\pi_j = 1 - \widehat{f}_j(0)$.
- A severity distribution $\widehat{h}_j \in \mathscr{P}_{11}$ which is given by

$$\widehat{h}_j(x) = \frac{\widehat{f}_j(x)}{\pi_j} \quad \text{for} \quad x = 1, 2, \ldots \tag{9.68}$$

Then calculate the De Pril transform via the new representation

$$\varphi_{f_j}(x) = -x \sum_{n=1}^{x} \frac{1}{n} \left(\frac{\pi_j}{\pi_j - 1} \right)^n \widehat{h}_j^{(n)*}(x), \quad x = 1, 2, \ldots \tag{9.69}$$

Step 2. Find the De Pril transform $\varphi_{f^{(n)*}(x)}$ of the convolved distribution $\widehat{f}^{(n)*}(x) = *_{j=0}^{n} \widehat{f}_j$ by simply summing the n De Pril transforms.

Step 3: Evaluate the density of the n-fold convolution using the result of Sundt & Vernic (2009) to obtain the density given in closed form by

$$\widehat{f}^{(n)*}(x) = -\frac{1}{x} \sum_{n=1}^{x} \widehat{f}^{(n)*}(x-n) \sum_{j=1}^{M} \left(\frac{\pi_j}{\pi_j - 1} \right)^n, \quad x = 1, 2, \ldots \tag{9.70}$$

There are also many approximation-based techniques for evaluation of the n-fold convolution and a detailed account of these can be found in Sundt & Vernic (2009, chapter 7).

Next we explore some special sub-families of distributions that have been discussed in previous chapters for their particular useful properties when utilized within an LDA model structure, these are the infinitely divisible distributional families. We discuss what special features the De Pril transform will have when applied to convolutions involving this family of distributions.

9.5.4 DE PRIL TRANSFORMS AND CONVOLUTIONS OF INFINITELY DIVISIBLE DISTRIBUTIONS

The De Pril transform is especially relevant in the context of infinitely divisible distributions, see details in Sundt & Vernic (2009, chapter 4, theorem 4.1, 4.2 and 4.5 and corollary 4.1). There computationally efficient properties that can be developed for the evaluation of the De Pril transform of such classes of distributions, as will be explored in the next few results.

On the basis of the properties of the family of infinitely divisible distributions, one can then study the De Pril transform for such a class of distributions, including convolutions of such distributions. The first identity of interest is to recall the properties of compound Poisson distributions discussed in Chapter 6 in terms of convolutions of such distributions, see Cont & Tankov (2004). This result can clearly be seen as an efficient recursive procedure for the evaluation of aggregation of a number of independent single-loss LDA risk processes in financial hierarchical banking structure.

Proposition 9.2 (Convolutions of Compound Poisson Distributions: Multiple LDA Risks) *Consider m independent compound processes, given by annual loss random variables $Z^{(j)} = \sum_{n=0}^{N^{(j)}} X_n^{(j)} \sim F_{Z^{(j)}}$, each representing a single risk process and given by severity*

distributions $X^{(j)} \sim f_j$ *and a Poisson frequency distribution* $N^{(j)} \sim Poisson(\lambda_j)$ *for all* $j \in \{1, 2, \ldots, m\}$. *If one considers the convolution of each of these compound processes given by*

$$Z_T = \sum_{j=1}^{m} Z^{(j)} \sim F_{Z_T} = *_{j=1}^{m} F_{Z^{(j)}}, \tag{9.71}$$

where $F^{(m)*}$ *is a compound Poisson distribution with rate parameter in the frequency distribution* λ_T *and severity distribution* $f_T(x)$ *given by*

$$\lambda_T = \sum_{j=1}^{m} \lambda_j, \text{ and } f_T(x) = \frac{1}{\lambda_T} \sum_{j=1}^{m} \lambda_j f_j(x). \tag{9.72}$$

Now utilizing the properties of infinitely divisible distributions and their unique representation as compound Poisson distributions detailed in point 9 of Theorem 9.2 which states that an infinitely divisible distribution in the class $\mathscr{P}_{10}$ can always be expressed as a compound Poisson distribution with severity distribution in the class $\mathscr{P}_{11}$. Hence, if one considers the m-fold convolution of infinitely divisible discrete distributions $\widehat{f}_j \in \mathscr{P}_{10}$ for all $j \in \{1, 2, \ldots, m\}$, then one can utilize this representation for each distribution and re-express the solution as the m-fold convolution of compound Poisson distributions and utilize the results in Proposition 9.2. Alternatively, there is a different approach one may adopt in addressing this problem utilizing the De Pril transform, given in Proposition 9.3.

Proposition 9.3 (De Pril Transform and Infinitely Divisible Severity Distributions) *Consider the* m*-fold convolution of discrete infinitely divisible distributions* $\widehat{f}_j \in \mathscr{P}_{10}$ *for all* $j \in \{1, 2, \ldots, m\}$. *One could evaluate the compound Poisson distribution representation of each distribution*

$$\widehat{f}_j(x) = \sum_{n=0}^{\infty} \frac{\lambda_j^n}{n!} \exp(-\lambda_j) \widehat{g}_j(x), \tag{9.73}$$

which would involve determining for the Poisson frequency distribution, the λ_j *and the severity distribution* $\widehat{g}_j(x)$. *Then the application of Proposition 9.2 and Theorem 9.10 would produce a solution involving first the evaluation of the compound Poisson representation of each infinitely divisible distribution and then the* m*-fold infinitely divisible distribution with total intensity* λ_T *and severity* $\widehat{f}_T(x)$ *according to*

$$\begin{aligned}
\lambda_T &= \sum_{j=1}^{m} \lambda_j = -\sum_{j=1}^{m} \ln g_j(0), \\
\widehat{f}_T(x) &= \frac{1}{\lambda_T} \sum_{j=1}^{m} \lambda_j \widehat{f}_j(x) \\
&= \frac{1}{\lambda_T} \sum_{j=1}^{m} \lambda_j \left[\frac{1}{\widehat{g}_j(0)} \left(\frac{\widehat{g}_j(x)}{\lambda_j} - \frac{1}{x} \sum_{y=1}^{x-1} y \widehat{f}_j(y) \widehat{g}_j(x-y) \right) \right], \ \forall x \in \{1, 2, \ldots\}.
\end{aligned}$$

However, this computation can be performed in an alternative manner via the De Pril transform which avoids the need to evaluate each λ_j *and more importantly the recursions for evaluation of*

$\widehat{f}_j(x)$. *Instead knowledge of the De Pril transform of each distribution* f_j *is utilized to evaluate the De Pril transform of the* m*-fold convolution* $\widehat{f}^{(m)*}(x) = *_{j=1}^{m}\widehat{f}_j(x)$ *directly as follows:*

$$\varphi_{f^{(m)*}} = \sum_{j=1}^{m}\varphi_{f_j} = \sum_{j=1}^{m}\frac{1}{\widehat{f}_j(0)}\left(x\widehat{f}_j(x) - \sum_{y=1}^{x-1}\varphi_{f_j}(y)\widehat{f}_j(x-y)\right),\ \forall x \in \{1, 2, \ldots\},$$

and the resulting distribution is then the solution

$$\widehat{f}^{(m)*}(x) = \frac{1}{x}\sum_{y=1}^{x}\varphi_{f^{(m)*}(x)}(y)\widehat{f}^{(m)*}(x-y),\ \forall x \in \{1, 2, \ldots\}.$$

EXAMPLE 9.4 De Pril's Transform of a Poisson Distribution

Consider the Poisson distribution $\widehat{f}$ with parameter λ, then the De Pril transform is given trivially by considering the Poisson distribution as a compound Poisson distribution with a severity $\widehat{g}$ concentrated on unity, resulting in the expression

$$\varphi_f(y) = \lambda\mathbb{I}[y = 1]. \tag{9.74}$$

It is also often convenient in OpRisk LDA models to note the result in Theorem 9.13 which relates a mixed Poisson distribution to a compound distribution, which can be easily evaluated using the De Pril transform.

Theorem 9.13 (Mixed Poisson Distributions) *A mixed Poisson distribution that has a mixing distribution in the class* $\mathscr{P}_{10}$ *can be re-expressed according to a compound distribution with the mixing distribution making up the frequency distribution and a severity distribution given by a Poisson distribution with unit rate.*

There are several general extensions to such results provided in the actuarial literature, see, for example, details of the Wilmot class of mixing distributions in Sundt & Vernic (2009, chapter 3).

9.5.5 RECURSIONS FOR n-FOLD CONVOLUTIONS (PARTIAL SUM) DISTRIBUTION TAILS WITH DISCRETIZED SEVERITY

The motivation for this section was provided in Chapter 7 where it was seen that one can decompose heavy-tailed LDA distribution tails into a series of terms which correspond to transformations of tails of the n-fold convolution tails for $n = 1, 2, \ldots$. We recall that the higher order tail asymptotic expansions under Decomposition Approach 1 were obtained via the expression derived by Barbe & McCormick (2009a, section 2) and Albrecher *et al.* (2006) and according to

$$\overline{F^{(n)*}} = \sum_{1\leq k\leq n} T_{F,w}^{k-1}T_{F^{(n-k)*},1-w}\overline{F} + \sum_{1\leq k\leq n} T_{F,w}^{k-1}(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}), \tag{9.75}$$

where for distribution function $F(x)$ the three operators for convolution, scaling of the argument and differentiation for a measurable test function $\phi(x)$ are recalled to be given by

$$T_{F,w}\phi(x) = \int_0^{wx} \phi(x-\tau)\,dF(\tau), \text{ and}$$

$$M_c F(x) = \mathbb{P}\mathrm{r}[cX \le x] = \begin{cases} F\left(\frac{x}{c}\right), & \text{if } c > 0, \\ \overline{F}\left(\frac{x}{c^-}\right), & \text{if } c < 0, \\ \mathbb{I}[x \ge 0], & \text{if } c = 0. \end{cases} \tag{9.76}$$

Here the operator M_c when acting on distribution F provides the distribution for cX when c is positive, and for $c = 0$, one obtains the distribution function $M_0 F$ of the point mass at the origin. Note that for any $c > 0$, one can also observe the useful property that $\overline{M_c F}(x) = M_c \overline{F}(x)$.

Clearly, if one discretizes the severity distribution, then the challenge will be to explicitly calculate the terms given by the tails of the n-fold convolution $\overline{F^{(n)*}}$ and the distribution $F^{(n)*}$. The recursions presented previously based on the De Pril transforms can be extended to allow one to develop recursions which are numerically robust directly for the tail of the n-fold distributions.

In Sundt & Vernic (2009, chapter 8), a class of recursions for cumulative distributions as well as tail functions are presented in a very general framework that generalizes the De Pril transforms above to the class $\mathscr{F}_1$ of unnormalized distributions. In this chapter, we restrict attention to a small subset of these recursions primarily for evaluation of the terms $\overline{F^{(n)*}}$ and $F^{(n)*}$.

The first result is given in Theorem 9.14 and shows the recursion for the distribution function and the tail function of a discretized distribution, see Sundt & Vernic (2009, theorem 8.1).

Theorem 9.14 (Recursions for Discretized Loss Model Distribution and Tails) *Consider a distribution $f \in \mathscr{F}_{10}$ that satisfies the relationship given by the recursion*

$$\widehat{f}(x) = q(x) + \sum_{i=1}^{k} \left(a(i) + \frac{b(i)}{n} \right) \widehat{f}(x-i), \tag{9.77}$$

with $q(x) = \widehat{f}(x) - \sum_{i=1}^{k} \left(a(i) + \frac{b(i)}{n} \right) \widehat{f}(x-i)$ and $k \in \mathbb{J}^+$. Then the discretized distribution and tail of the distribution are given by the recursions:

$$\widehat{F}(x) = q(x) + \sum_{i=1}^{k+1} \left(a(i) - a(i-1) + \frac{b(i) - b(i-1)}{n} \right) \widehat{F}(x-i), \quad n = 1, 2, \ldots,$$

$$\overline{\widehat{F}}(x) = -q(x) + \sum_{i=1}^{k+1} \left(a(i) - a(i-1) + \frac{b(i) - b(i-1)}{n} \right) \overline{\widehat{F}}(x-i), \quad n = 1, 2, \ldots,$$

with $a(0) = -1$ and initial value

$$\overline{\widehat{F}}(n) = \sum_{i=n+1}^{\infty} \overline{\widehat{F}}(i), \quad n = 0, -1, \ldots, 1-k-1.$$

The recursion for the distribution can also be written as a generalization of the De Pril transform as given in Theorem 9.15, see Sundt & Vernic (2009, corollary 8.6).

Theorem 9.15 (Generalized De Pril Transform for Distributions) *Consider a discretized distribution $\widehat{f} \in \mathscr{P}_{10}$, then the following generalized De Pril transform applies for the cumulative distribution*

$$\widehat{F}(x) = \frac{1}{x}\sum_{y=1}^{x}(\varphi_f(y)+1)\widehat{F}(x-y), \quad x = 1, 2, \ldots \tag{9.78}$$

These results provide recursions for a discretized loss model; however, to evaluate the quantities of interest in this section, these recursions should be extended to the n-fold convolution setting. Therefore, the second set of recursions that will be of relevance to this section will be for the cumulative distribution and tail of the n-fold convolutions given by $\overline{\widehat{F}^{(n)*}}$ and $\widehat{F}^{(n)*}$ as shown in Theorem 9.16, see details in Sundt & Vernic (2009, corollary 8.2).

Theorem 9.16 (Recursions for Discretized Loss Model n-Fold Distribution) *Consider a distribution $\widehat{f} \in \mathscr{P}_{10}$, then for any $n \in \mathbb{J}^+$, one has the recursion for the discretized distribution $\widehat{F}(x)$ given by*

$$\widehat{F}^{(n)*}(x) = \frac{1}{\widehat{f}(0)}\sum_{i=1}^{x}\left[\left(\frac{i}{x}(M+1)-1\right)\widehat{f}(x)+\frac{1}{x}\widehat{F}(i-1)\right]\widehat{F}^{(n)*}(x-i),$$
$$M = 1, 2, \ldots, \quad x = 1, 2, \ldots$$

9.6 Estimating Higher Order Tail Approximations for Convolutions with Continuous Severity Distributions (Fixed n)

In this section, we consider how to remove the restriction imposed above of discretization of the severity distribution when performing the evaluation of the n-fold convolution in a recursive fashion. This is because in heavy-tailed compound process severity models, the discretization grid range and resolution can easily become computationally prohibitive. It is, therefore, of direct interest to both reduce the computational cost associated with a large numbers of grid point required for sub-exponential models and improve the computational efficiency and accuracy by developing efficient Monte Carlo-based approximation schemes by looking at the solutions from a statistical perspective rather than a deterministic numerical integration perspective. In this section, we detail the continuous distribution recursive relationships which one may consider, and in Section 9.8, we detail the Monte Carlo sampling procedures.

Therefore, in the following sections, we particularly focus on integration of advanced Monte Carlo-based methodologies which when coupled with more standard actuarial recursions for the evaluation of a compound sum distribution or density at a given point (quantile level) can efficiently evaluate desired quantities for OpRisk practitioners. In particular, we provide a means of performing estimation for the evaluation of the following quantities for

distribution and densities of the n-fold convolutions which are of direct interest to OpRisk measurement:

$$F_{Z_n}(x) = \sum_{s=0}^{n} p_s F^{(s)*}(x) \text{ and } f_{Z_n}(x) = \sum_{s=0}^{n} p_s f^{(s)*}(x),$$
$$\overline{F}_{Z_n}(x) = \sum_{s=1}^{n} p_s \overline{F^{(s)*}}(x).$$

Under the approaches developed, it will be demonstrated how to efficiently quantify several important quantities related to the rate of decrease of bias and variance (as a function of quantile level) that quantify the accuracy of the asymptotic closed-form approximations (for first-order, second-order and higher order schemes) for the tails of a compound sum. This will, therefore, provide a means of assessing the accuracy and suitability of a closed-form tail approximation or ultimately a risk measure approximation, for given finite quantile level as presented in detail in Chapter 8.

In Chapter 7 in Proposition 7.28, it was demonstrated that one can obtain the following *exact* recursive decomposition of the tail of the n-fold convolution for the distribution and density, where given n i.i.d. losses, the tail distribution of the resulting convolved distribution $\overline{F^{(n)*}}(x)$ can be decomposed as follows for any $w \in (0,1)$ via

$$\overline{F^{(n)*}} = \sum_{1 \le k \le n} T_{F,w}^{k-1} T_{F^{(n-k)*},1-w} \overline{F} + \sum_{1 \le k \le n} T_{F,w}^{k-1} (M_{1/(1-w)} \overline{F^{(n-k)*}} M_{1/w} \overline{F}).$$

Therefore, one can also define the density for the n-fold convolution which will be given by differentiation and can be easily shown to result in the recursion in Proposition 9.4, see Barbe & McCormick (2009b, proposition 5.1.1 and section 5.2).

Proposition 9.4 (Recursive Representation of n-fold Convolved Density) *Given n i.i.d. losses, the tail distribution of the resulting convolved distribution $\overline{F^{(n)*}}(x)$ can be decomposed as follows for any $w \in (0,1)$ via*

$$\overline{F^{(n)*}} = \sum_{1 \le k \le n} T_{F,w}^{k-1} T_{F^{(n-k)*},1-w} \overline{F} + \sum_{1 \le k \le n} T_{F,w}^{k-1} (M_{1/(1-w)} \overline{F^{(n-k)*}} M_{1/w} \overline{F}).$$

Utilizing this recursion for the tail of the distribution, one can also derive through the identities in Proposition 7.26 the following recursion for the k-th derivative of the n-fold tail convolution distribution.

$$\frac{d^k}{dx^k}\overline{F^{(n)*}}(x) = T_{F,w}\frac{d^k}{dx^k}\overline{F^{(n-1)*}}(x) + T_{F^{(n-1)*},1-w}\frac{d^k}{dx^k}\overline{F}(x) - \sum_{1 \le i \le k-1} M_{1/w}\left[\frac{d^i}{dx^i}\overline{F}\right] M_{1/(1-w)}\frac{d^{(k-i)}}{dx^{(k-i)}}\overline{F^{(n-1)*}}(x).$$

Furthermore, if the hazard function $h(x) = F^{(1)}(x)/\overline{F}(x)$ exists for the chosen severity distribution and one assumes that the severity distribution $F(x)$ has no atoms in $[0,\infty)$, then the n-fold convolution of such a severity distribution will also admit no atoms in the support $[0,\infty)$. In addition, the continuity and boundedness of the severity distribution $F(x)$ is preserved under an n-fold convolution. Therefore, if the severity distribution $F(x)$ admits a density $f(x) = F^{(1)}(x)$, then

so does the distribution of the n-fold convolution $f^{(n)*}(x) = dF^{(n)*}(x)/dx$. Hence, obtaining the recursion for the n-fold convolved density can be achieved via the tail distribution expansion according to $\frac{d}{dx}\overline{F^{(n)*}}(x)$ which simply involves selecting $k = 1$ to obtain

$$\frac{d}{dx}\overline{F^{(n)*}}(x) = \frac{d}{dx}(1 - F^{(n)*}(x)) = -\frac{d}{dx}F^{(n)*}(x) = -f^{(n)*}(x);$$

therefore, we see that the density can be represented by

$$\begin{aligned} f^{(n)*}(x) &= -\frac{d}{dx}\overline{F^{(n)*}}(x) \\ &= -T_{F,w}\frac{d}{dx}\overline{F^{(n-1)*}}(x) - T_{F^{(n-1)*},1-w}\frac{d}{dx}\overline{F}(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{F^{(n-1)*},1-w}f(x). \end{aligned}$$

From this decomposition, we note the following properties of such a recursive decomposition which will be advantageous to the class of Monte Carlo-based sampling strategies one can develop for the evaluation of such quantities in practice.

Remark 9.8 *If one were to calculate the tail functional for some large quantile (i.e. $x \gg 0$) in a standard manner, it would require integration given by the following expression:*

$$\overline{F^{(n)*}}(x) = \int_x^\infty f^{(n)*}(x)\,dx = 1 - \int_0^x f^{(n)*}(x)\,dx, \tag{9.79}$$

which would be difficult both from the perspective that it requires evaluation of the n-fold convolution density $f^{(n)*}(x)$ and, in addition, it requires the domain of integration to be either deep into the tails of this density $[x, \infty)$ or accurate over the domain $[0, x)$. Therefore, in either case, one requires to work deep in the tails of the integration, which would typically require rare-event simulations. Therefore, one advantage of the representation developed is that it requires only integration over the domains (selected by user choice of $w \in [0, 1]$) for the first term, where assuming the density of the severity model exists, will be given by

$$T_{F,w}f^{(n-1)*}(x) = \int_0^{wx} f^{(n-1)*}(x-\tau)\,dF(\tau) = \int_0^{wx} f^{(n-1)*}(x-\tau)f(\tau)d\tau, \tag{9.80}$$

and for the second term given by

$$T_{F^{(n-1)*},1-w}f(x) = \int_0^{(1-w)x} f(x-\tau)\,dF^{(n-1)*}(\tau) = \int_0^{(1-w)x} f(x-\tau)f^{(n-1)*}(\tau)d\tau. \tag{9.81}$$

Therefore, if one selected $w = 0.5$, then the integration domain is halved, thereby improving the ability to accurately estimate these integrals to evaluate $\overline{F^{(n)*}}(x)$ for large quantiles, as the region in which samples are required from the tails of the distribution is reduced, this will be beneficial.

Now, in OpRisk, it is important to be able to evaluate the distribution $\overline{F^{(n)*}}(x)$ and density $f^{(n)*}(x)$ for large quantile values, that is, values of $x \gg 0$. What was demonstrated in Chapter 6 was how to achieve this evaluation analytically with closed-form representations as

$x \to \infty$. These closed-form representations involved numerous stages of approximation and assumptions relating to the severity distribution $F(x)$ tail attributes. In this section, we recall each stage of approximation and we develop numerical procedures which efficiently allow us to quantify the accuracy of these approximations for finite quantile levels and to assess the rate of convergence of such approximations. This will then provide a valuable tool for practitioners to either utilize an efficient Monte Carlo-based tail estimation or assess the accuracy (bias and variance) of such a closed-form asymptotic approximation and the rate of convergence of such asymptotic approximations. We will focus here on Decomposition Approach 1 in Chapter 7.

To proceed, we note that an *o*-estimate provides a qualitative statement, where $f(x) = o(g(x))$ tells one that the quotient $f(x)/g(x)$ tends to 0 as $x \to \infty$. However, this statement is not quantitative as it does not provide information on the rate of this convergence; it is this point that is important to study and quantify if one also wishes to make use of such asymptotic closed-form approximations. Unfortunately, obtaining closed-form expressions for the rate is a very challenging task that remains an open problem for the majority of approximations obtained. Therefore, we provide in this section an efficient Monte Carlo methodology to assess the rate of such approximations that is not context specific.

In cases in which an *O*-estimate is known for the order of asymptotic approximation, one has a more precise functional characterization of the behaviour $f(x) = O(g(x)/h(x))$ with some explicit function $h(x)$ that tends to infinity as $x \to \infty$ implies the *o*-estimate $f(x) = o(g(x))$ and provides more information. Although more informative, this often does not provide a tight enough rate of convergence for practical purposes; it is, therefore, the intention of this section to assess such properties. Unfortunately, the majority of results developed to date for the series representation of the tail of an n-fold distribution are only attained as little $o(\cdot)$ as summarized below, making it particularly important to understand the suitability of such approximations and efficient algorithms to help quantify the variance, bias and rates of convergence of such estimations as $x \to \infty$. In addition, it will aid in the assessment of the contribution of additional terms in higher order asymptotic series expansions of tail functionals when undertaking closed-form representations for risk measures.

9.6.1 APPROXIMATION STAGES TO BE STUDIED

Here we provide a summary of the approximation stages undertaken in the construction of the closed-form higher order asymptotic tail expansion results for the n-fold distributional convolutions.

1. *Stage 1: Recursions for Distribution and Density.* Common to every stage of the analysis of the tail expansion for the n-fold convolutions of the distribution and density functions is recursions for the representation of the density or distribution. Here we focus on four recursive relationships that could be adopted, two of which are exact and the other two are based on asymptotic approximations derived from particular features of the heavy-tailed severity distributions considered.
   (a) *Recursive density representation 1 (exact)* is suitable for all severity distributions and given by the standard n-fold convolution representation for a density
   $$f^{(n)*}(x) = \int_0^\infty f^{(n-1)*}(x-\tau) f(\tau) d\tau.$$
   (b) *Recursive density representation 2 (exact)* is suitable for all severity distributions and obtained through differentiation of the n-fold convolved tail of the severity model's

decomposition according to

$$\begin{aligned} f^{(n)*}(x) &= -\frac{d}{dx}\overline{F^{(n)*}}(x) = -T_{F,w}\frac{d}{dx}\overline{F^{(n-1)*}}(x) - T_{F^{(n-1)*},1-w}\frac{d}{dx}\overline{F}(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{F^{(n-1)*},1-w}f(x). \end{aligned}$$

(c) *Recursive density representation 3 (approximate)* is suitable for the class of sub-exponential severity distributions and given for large $x \gg 0$ by approximation

$$\begin{aligned} f^{(n)*}(x) &= -\frac{d}{dx}\overline{F^{(n)*}}(x) = -T_{F,w}\frac{d}{dx}\overline{F^{(n-1)*}}(x) - T_{F^{(n-1)*},1-w}\frac{d}{dx}\overline{F}(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{F^{(n-1)*},1-w}f(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{1-\overline{F^{(n-1)*}},1-w}f(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{1-(n-1)\overline{F},1-w}f(x) \\ &= T_{F,w}f^{(n-1)*}(x) + T_{\widetilde{F},1-w}f(x), \end{aligned}$$

where $\widetilde{F}(x) = F(x)(1-n) - n$.

(d) *Recursive density representation 4 (approximate)*. If the severity distribution is assumed to satisfy the assumptions in Proposition 7.27, then one may consider the second representation obtained by differentiating the asymptotic approximate expression after taking out the leading terms from the big O-notation and assuming continuity of the severity distribution so as to swap the order of integration and differentiation to obtain

$$\begin{aligned} f^{(n)*}(x) &= -\frac{d}{dx}\overline{F^{(n)*}}(x) \\ &= -\frac{d}{dx}\left[\frac{n}{n-1}\int_0^{x/2}\overline{F^{(n-1)*}}(x-\tau)\,dF(\tau) + O\left(\overline{F}\Big(\frac{x}{2}\Big)^{2-\varepsilon}\right)\right] \\ &= -\frac{d}{dx}\left[\frac{n}{n-1}T_{F,1/2}\overline{F^{(n-1)*}}(x) + \overline{F}\Big(\frac{x}{2}\Big)^{2-\varepsilon}O(1)\right] \\ &= \frac{n}{n-1}T_{F,1/2}f^{(n-1)*}(x) + (1-\varepsilon/2)\overline{F}\Big(\frac{x}{2}\Big)^{1-\varepsilon}f\left(\frac{x}{2}\right)O(1) \\ &= \frac{n}{n-1}T_{F,1/2}f^{(n-1)*}(x) + O\left(\overline{F}\Big(\frac{x}{2}\Big)^{1-\varepsilon}f\left(\frac{x}{2}\right)\right) \\ &= \frac{n}{n-1}T_{F,1/2}f^{(n-1)*}(x). \end{aligned} \tag{9.82}$$

2. *Stage 2: Truncation of the Taylor Series Expansion of n-fold Distribution Tail.* Under 'Decomposition Approach 1' which produces the recursive decomposition of the tail distribution of n i.i.d. losses for any $w \in (0,1)$ according to

$$\overline{F^{(n)*}}(x) = \sum_{1\le k\le n} T_{F,w}^{k-1}T_{F^{(n-k)*},1-w}\overline{F}(x) + \sum_{1\le k\le n} T_{F,w}^{k-1}(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F})(x).$$

Then one may obtain from this expansion a Taylor series representation which would be truncated after a desired number of terms, leaving a remainder term. Assuming that

the severity distribution tail $\overline{F}$ is differentiable up to a positive integer order $m+1 \in \mathbb{J}$, then the mth order Taylor series representation about a point x_0 is given for each fixed $k \in \{1, \ldots, n\}$ for a value of $\kappa \in \{w, (1-w)\}$ according to the two sets of terms:

$$T_{F^{(n-k)*},\kappa}\overline{F}(x) \approx \sum_{0 \le j \le m} \frac{(-1)^j \overline{F}^{(j)}(x)}{j!} \int_0^{\kappa x} \tau^j dF^{(n-k)*}(\tau), \quad \text{for } x = x_0 \in [0, \infty),$$

and for the second set of recursive terms, one has the following expressions for each $k \in \{1, \ldots, n\}$ given by

$$\begin{aligned} &T_{F,w}^{k-1}(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F})(x) \\ &\quad = T_{F,w}^{k-2} \int_0^{wx} \overline{F^{(n-k)*}}((1-w)x - \tau_1)\overline{F}(wx - \tau_1)dF(\tau_1). \end{aligned}$$

For each of these terms, there is then a set of remainder terms. In the case of the first terms, for each fixed $k \in \{1, \ldots, n\}$ and a value of $\kappa \in \{w, (1-w)\}$, the remainder terms were given analytically for any $x_* \in [x_0, x]$ by

$$T_{F^{(n-k)*},\kappa}R_m(x) = \frac{(-1)^{m+1}\overline{F}^{(m+1)}(x)}{(m+1)!} \int_0^{\kappa x} \tau^{m+1} dF^{(n-k)*}(\tau), \quad \text{for } x = x_0 \in [0, \infty).$$

The remainder terms can be obtained analogously for this second set of terms by substitution of the series expansion remainder for $\overline{F}$ into each term. However, in general, in Proposition 7.31, it was shown that one obtains an asymptotic order of approximation for the remaining terms under an assumption that the severity distribution has a tail distribution that is regularly varying with $\overline{F}(x) \in RV_{-\rho}$. In this case, the remainder of the series expansion of $\overline{F}$ is given by $R_m(x) \sim o(\overline{F}^{(m)}(x))$ which in turn results in the first set of terms $T_{F^{(n-k)*},\kappa}R_m(x)$ being of the same asymptotic order of approximation for all $k \le n$, that is, $T_{F^{(n-k)*},\kappa}R_m(x) \sim o(\overline{F}^{(m)}(x))$. If one further assumes that the hazard rate $h(x)$ exists, for sub-exponential distribution F, then from the second set of terms in the decomposition, the asymptotic behaviour for each of the terms for $k \in \{1, 2, \ldots, n\}$ was shown to satisfy the following bound on interval $[t, \infty)$ with a positive number ε and constant A such that for any positive $k \le n$,

$$|T_{F,w}^{k-1}(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F})| \le A(1+\varepsilon)^n h^{m+1}\overline{F}. \tag{9.83}$$

Furthermore, the order of approximation of these terms for any $k \le n$ can be shown to be

$$T_{F,w}^{k-1}(M_{1/(1-w)}\overline{F^{(n-k)*}}M_{1/w}\overline{F}) = o(\overline{F}^{(m)}(x)). \tag{9.84}$$

3. *Stage 3: Modified Taylor Series Expansion of n-fold Distribution Tail.* Under 'Decomposition Approach 1', it was shown in Propositions 7.31 and 7.31 that a modified series expansion was developed in which it was demonstrated that for all $j \le m$ and $k \le n$, the terms given by

$$\overline{F}^{(j)}(x) \int_{\tau x}^{\infty} \tau^j dF^{(k)*}(\tau) = o(\overline{F}^{(m)}(x)), \tag{9.85}$$

have the same order of approximation as the remainder terms, because of the assumption of a monotone density and Karamata's theorem. This resulted in the mth order Taylor

series representation about a point $x = x_0$ given asymptotically in $x \to \infty$ up to order of approximation $o(\overline{F}^{(m)}(x))$ by the modified Taylor series

$$\overline{F^{(n)*}}(x) = \sum_{1 \le k \le n} T_{F,w}^{k-1} T_{F^{(n-k)*},1-w} \overline{F}(x) + \sum_{1 \le k \le n} T_{F,w}^{k-1} (M_{1/(1-w)} \overline{F^{(n-k)*}} M_{1/w} \overline{F})(x)$$

$$= \sum_{1 \le k \le n} \sum_{0 \le j \le m} T_{F,w}^{k-1} \frac{(-1)^j \mu_{F^{(n-k)*},j}}{j!} \overline{F}^{(j)}(x).$$

Remark 9.9 *From a statistical perspective, to assess each of these stages of approximation, it is clear that one requires a means of approximation of these integrals. To achieve this, we will consider a Monte Carlo-based solution which utilizes sequential Monte Carlo (SMC) that will allow us to obtain estimation of each term in the tail decompositions. In particular, we will discuss later the key challenge in the above sequence of steps to perform the tail asymptotic approximation, which involves the sampling of the sequence of distributions* $\{f_{Z_k}(x)\}_{k=1}^n = \{f_X^{(k)*}(x)\}_{k=1}^n$.

The sampling procedures considered here will be based on a class of numerical interacting particle methods known in statistics as SMC samplers algorithms, which we describe in detail next.

9.7 Sequential Monte Carlo Sampler Methodology and Components

Consider sampling from a sequence of probability distributions denoted generically by $\{\pi_t(d\boldsymbol{\theta})\}_{t=1}^T$, where in the context of this section, we could be considering quantities such as those in the following list of examples:

- *Example 1*. Consider a sequence of densities of interest given by $\{\pi_t(x)\}_{t=1}^T$ where each density π_t in the sequence is given by

$$\pi_t(x) := f_X^{(t)*}(x) = \int_0^\infty f_X^{(t-1)*}(x-\tau) f_X(\tau) d\tau. \tag{9.86}$$

- *Example 2*. Consider a sequence of densities of interest given by $\{\pi_t(x|x > q)\}_{t=1}^T$ where each density π_t in the sequence is given by

$$\pi_t(x) := f_X^{(t)*}\mathbb{I}[x > q] = \int_q^\infty f_X^{(t-1)*}(x-\tau) f_X(\tau) d\tau, \tag{9.87}$$

  for some quantity $q > 0$.

- *Example 3*. Consider a sequence of densities of interest corresponding to an aggregate sum $x_D = \sum_{i=1}^d X_i$ of random variables (such as an aggregate of several loss processes) given by $\left\{\pi_t(x_D|\sum_{i=1}^d X_i > q_t)\right\}_{t=1}^T$ where each density π_t in the sequence is given by

$$\pi_t(x) := *_{i=1}^d f_{X_i}(x)\mathbb{I}\left[\sum_{i=1}^d > q_t\right], \tag{9.88}$$

  for some sequence of level sets (quantiles) given by $0 < q_1 < q_2 < \cdots < q_T$.

There are many other examples for sequences of distributions that may be of interest in the OpRisk setting, the above are just a few illustrations of relevance to this section of the book.

In all such examples, the general aim of the SMC sampler framework will be to develop a large collection of N weighted random samples at each time t denoted by $\{W_t^{(i)}, \boldsymbol{\Theta}_t^{(i)}\}_{i=1}^N$ such that $W_t^{(i)} > 0$ and $\sum_{i=1}^N W_t^{(i)} = 1$ which represent a discrete 'random' particle grid weighted approximation of the target distribution. These importance weights and samples, $\{W_t^{(i)}, \boldsymbol{\Theta}_t^{(i)}\}_{i=1}^N$, are known as particles (hence, the name often given to such algorithms as particle filters or interacting particle systems). For such approaches to be sensible, we would require that the empirical distributions constructed through these samples should converge asymptotically ($N \to \infty$) to the target distribution π_t for each time t. This means that for any π_t integrable function, say denoted by $\phi(\boldsymbol{\theta}) : E \to \mathbb{R}$, one would have the following convergence:

$$\sum_{i=1}^{N} W_t^{(i)} \phi(\boldsymbol{\theta}_t^{(i)}) \xrightarrow{a.s.} \mathbb{E}_{\pi_t}[\phi(\boldsymbol{\Theta})]. \tag{9.89}$$

The sequential nature of such algorithms arises from the fact that they iteratively construct the sets of weighted particles recursively through a sequential importance sampling framework, see many examples of such algorithms in Doucet *et al.* (2000), Oh & Berger (1993), Givens & Raftery (1996), Gilks & Berzuini (2002), Neal (2001) and the tutorial of Doucet & Johansen (2009).

In the following SMC samplers algorithm, a particular variant of SMC algorithms, a modification of the sequential Monte Carlo algorithm is developed. Consider a generic sequence of distributions given by $\pi_t(\boldsymbol{\theta}), t = 1, \ldots, T$, with $\boldsymbol{\theta} \in E$, where the final distribution π_T is the distribution of interest. By introducing a sequence of backward kernels L_k, a new distribution

$$\widetilde{\pi}_t(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t) = \pi_t(\boldsymbol{\theta}_t) \prod_{k=1}^{t-1} L_k(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k), \tag{9.90}$$

may be defined for the *path* of a particle $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t) \in E^t$ through the sequence $\pi_1, \ldots, \pi_t$. The only restriction on the backward kernels is that the correct marginal distributions $\int \widetilde{\pi}_t(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t) d\boldsymbol{\theta}_1, \ldots, d\boldsymbol{\theta}_{t-1} = \pi_t(\boldsymbol{\theta}_t)$ are available. Within this framework, one may then work with the constructed sequence of distributions, $\widetilde{\pi}_t$, under the standard SMC algorithm. In summary, the SMC sampler algorithm involves three stages:

1. *mutation*, whereby the particles are moved from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$ via a mutation kernel $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$;
2. *correction*, where the particles are reweighted with respect to π_t via the incremental importance weight (Equation 9.91); and
3. *selection*, where according to some measure of particle diversity, commonly the effective sample size, the weighted particles may be resampled in order to reduce the variability of the importance weights.

In more detail, suppose that at time $t-1$, the distribution $\widetilde{\pi}_{t-1}$ can be approximated empirically by $\widetilde{\pi}_{t-1}^N$ using N weighted particles. These particles are first propagated to the next distribution $\widetilde{\pi}_t$ using a mutation kernel $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ and then assigned new weights

$W_t = W_{t-1} w_t(\boldsymbol{\theta}_1, \ldots \boldsymbol{\theta}_t)$, where W_{t-1} is the weight of a particle at time $t-1$ and w_t is the incremental importance weight given by

$$w_t(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t) = \frac{\widetilde{\pi}_t(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t)}{\widetilde{\pi}_{t-1}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)} = \frac{\pi_t(\boldsymbol{\theta}_t) L_{t-1}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}. \tag{9.91}$$

The resulting particles are now weighted samples from $\widetilde{\pi}_t$. Consequently, from (Equation 9.91), under the SMC sampler framework, one may work directly with the marginal distributions $\pi_t(\boldsymbol{\theta}_t)$ such that $w_t(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_t) = w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$. While the choice of the backward kernels L_{t-1} is essentially arbitrary, their specification can strongly affect the performance of the algorithm, as will be discussed in the following subsections.

The basic version of the SMC sampler algorithm, therefore, proceeds explicitly as follows in Algorithm 9.3.

Remark 9.10 *In all cases in which we utilize the incremental importance sampling weight correction, the arguments in the expressions only need to be known up to normalization. That is, it is perfectly acceptable to only be able to evaluate the sequence of target distributions $\{\pi_t\}$ up to normalization constant. This is true as long as the same normalization constant is present for all particles, because the renormalization step will correct for this lack of knowledge in the importance weighting. In practice, this is critical to the application of such methods.*

Algorithm 9.3 (Sequential Monte Carlo Sampler)

1. *Initialize the particle system.*
   (a) *Set $n = 1$.*
   (b) *For $i = 1, \ldots, N$, draw initial particles $\boldsymbol{\Theta}_1^{(i)} \sim p(\boldsymbol{\theta})$.*
   (c) *Evaluate incremental importance weights $\{w_1(\boldsymbol{\Theta}_1^{(i)})\}$ using Equation 9.91 and normalize the weights to obtain $\{W_1^{(i)}\}$.*

   Iterate the steps below through each distribution in sequence $\{\pi_t\}_{t=2}^T$.
2. *Resampling*
   (a) *If the effective sampling size $ESS = \frac{1}{\sum_{i=1}^N (w_t^{(i)})^2} < N_{\text{eff}}$ is less than a threshold N_{eff}, then resample the particles via the empirical distribution of the weighted sample either by multinomial or stratified methods, see discussion on unbiased resampling schemes in Künsch (2005) and Del Moral (2004).*
3. *Mutation and Correction*
   (a) *Set $t = t + 1$, if $t = T + 1$, then stop.*
   (b) *For $i = 1, \ldots, N$, draw samples from mutation kernel $\boldsymbol{\Theta}_t^{(i)} \sim M_t(\boldsymbol{\Theta}_{t-1}^{(i)}, \cdot)$.*
   (c) *Evaluate incremental importance weights $\{w_1(\boldsymbol{\Theta}_1^{(i)})\}$ using Equation 9.91 and normalize the weights to obtain $\{W_1^{(i)}\}$ via*

$$W_t^{(i)} = W_{t-1}^{(i)} \frac{w_t^{(i)}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_t)}{\sum_{j=1}^N W_{t-1}^{(i)} w_t^{(i)}(\boldsymbol{\Theta}_{t-1}, \boldsymbol{\Theta}_t)}. \tag{9.92}$$

9.7.1 CHOICE OF MUTATION KERNEL AND BACKWARD KERNEL

There are many choices for mutation kernel and backward kernel that could be considered when designing an SMC sampler algorithm. In this section, we survey a few possible choices and note an important difference between the SMC sampler and Markov chain Monte Carlo (MCMC) methods in the following remark.

Remark 9.11 *In the MCMC methods presented previously, the proposal kernel was typically selected to ensure the resulting Markov chain satisfied reversibility and detailed balance conditions, or in the case of the adaptive proposals some notion of eventual non-adaption (diminishing adaptation and bounded convergence). Unlike the MCMC methods, in the case of the SMC samplers algorithms, the mutation kernel is significantly more flexible with regard to choice and adaption strategies. It is clear that the optimal choice of mutation kernel would be the next distribution in the sequence* $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \pi_t(\boldsymbol{\theta}_t)$*, as this would minimize the variance of the incremental weights, although clearly the context of the application of SMC samplers is one in which the target distribution cannot be sampled directly via inversion or rejection sampling methods, so this choice is not practical.*

Some examples of possible choices of the mutation kernel are given below:

1. **Independent Kernels.** In this setting, one would select a mutation kernel given for all $t \in \{1, 2, \ldots, T\}$ by $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = M_t(\boldsymbol{\theta}_t)$.
2. **Local Random Walks.** In this setting, the kernel would be selected for all $t \in \{1, 2, \ldots, T\}$ to be of the form $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ where the mutation from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\theta}_t$ follows a local random walk based around say a Gaussian smoothing kernel as in Givens & Raftery (1996).
3. **Markov chain Monte Carlo Kernels.** In this setting, the kernel would be selected for all $t \in \{1, 2, \ldots, T\}$ to be an MCMC kernel of invariant distribution π_t. As noted in Del Moral *et al.* (2006) and Peters (2005), this option is suitable if the Markov chain kernel is mixing rapidly or if the sequence of distributions is such that π_{t-1} is close to π_t, which is often the case by design. Then the use of an MCMC kernel would result in running for each stage, N inhomogeneous Markov chains. Then one must correct for the fact that one is not targeting the correct distribution under these Markov chains, which is achieved using importance sampling: $\widehat{\pi}_{t-1}^N = \sum_{i=1}^N W_{t-1}^{(i)} \delta_{\boldsymbol{\theta}_{t-1}^{(i)}}(\boldsymbol{\theta})$ and running L iterations of the Markov chain for each particle, where each of the N chains will target $\sum_{i=1}^N W_{t-1}^{(i)} \prod_{l=1}^L M_l(\boldsymbol{\theta}_{l-1}^{(i)}, \boldsymbol{\theta}_l)$ which is not in general π_t, then with an importance sampling correction such an approach is accurate and unbiased (i.e. targets the distribution of interest at time t given by π_t).
4. **Gibbs Sampler Kernels.** If the sequence of target distributions $\{\pi_t\}_{t \geq 0}$ is such that its support is multivariate, then it may also be possible to sample from the full conditional distributions in the sequence of distributions. This approach allows one to undertake a Gibbs step; this would involve a kernel for update of the kth element given in the form

$$M_t(\boldsymbol{\theta}_{t-1}, d\boldsymbol{\theta}_t) = \delta_{\boldsymbol{\theta}_{t-1,-k}}(d\boldsymbol{\theta}_{t,-k}) \pi_t(\boldsymbol{\theta}_{t,k} | \boldsymbol{\theta}_{t,-k}) \tag{9.93}$$

with $\boldsymbol{\theta}_{t,-k} = (\boldsymbol{\theta}_{t,1}, \boldsymbol{\theta}_{t,2}, \ldots, \boldsymbol{\theta}_{t,k-1}, \boldsymbol{\theta}_{t,k+1}, \ldots, \boldsymbol{\theta}_{t,J})$, where there are J parameters in the OpRisk model target posterior. If the full conditionals are not available, one could approximate them accurately at each stage and then correct for the approximation error through importance sampling.

5. **Mixture Kernels.** It is always possible to consider a mixture kernel choice given by

$$M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \sum_{m=1}^{M} \alpha_{t,m}(\boldsymbol{\theta}_{t-1}) M_{t,m}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t), \tag{9.94}$$

with $\alpha_{t,m}(\boldsymbol{\theta}_{t-1}) > 0$ and $\sum_{m=1}^{M} \alpha_{t,m}(\boldsymbol{\theta}_{t-1}) = 1$. One special case of this type of kernel would be an independent kernel constructed by a kernel density estimate of $M_{t,m}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ for all m and $\alpha_{t,m}(\boldsymbol{\theta}_{t-1}) = W_{t-1}^{(i)}$ with $M = N$.

6. **Partial Rejection Control Kernels.** In this case, one aims to construct a mutation kernel in the SMC sampler that guarantees all sampled particles have importance weights with a 'fitness' exceeding a user-specified threshold at each time t, denoted by c_t, such that $w_t^{(i)} \geq c_t, \ \forall i \in \{1, 2, \ldots, N\}$. To achieve this, one modifies any of the above mutation kernels to take a form given by

$$M_t^*(\boldsymbol{\theta}_{t-1}^i, \boldsymbol{\theta}_t) = r(c_t, \boldsymbol{\theta}_{t-1}^{(i)})^{-1} \left[\min \left\{ 1, W_{t-1}^{(i)} \frac{w_t\left(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t\right)}{c_t} \right\} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t) \right]. \tag{9.95}$$

The quantity $r(c_t, \boldsymbol{\theta}_{t-1}^{(i)})$ denotes the normalizing constant for particle $\boldsymbol{\theta}_{t-1}^{(i)}$, given by

$$r(c_t, \boldsymbol{\theta}_{t-1}^{(i)}) = \int \min \left\{ 1, W_{t-1}^{(i)} \frac{w_t\left(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t\right)}{c_t} \right\} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t) d\boldsymbol{\theta}_t. \tag{9.96}$$

Note that $0 < r(c_t, \boldsymbol{\theta}_{t-1}) \leq 1$ if (w.l.o.g.) the mutation kernel M_t is normalized, so that $\int M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) d\boldsymbol{\theta}_t = 1$, and if the partial rejection control (PRC) threshold $0 \leq c_t < \infty$ is finite. The sequence of PRC thresholds is then user specified to ensure a certain particle 'fitness' at each stage of the SMC sampler. We will detail more explicitly this example in a future section.

Proposition 9.5 (Optimal Backward Kernel) *Given any of the possible mutation kernels $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$, one can define the optimal backward kernel in the SMC sampler as the one that minimizes the variance of the incremental (unnormalized) importance sampling weights, given in Peters (2005) and Del Moral* et al. *(2006) by*

$$L_{t-1}^{opt}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \frac{\nu_{t-1}(\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{\nu_t(\boldsymbol{\theta}_t)}, \tag{9.97}$$

where one defines the sequence of integrated distributions on the path space by

$$\nu_t(\boldsymbol{\theta}_t) = \int \ldots \int \pi_1(\boldsymbol{\theta}_1) \prod_{l=1}^{t} M_l(\boldsymbol{\theta}_{l-1}, \boldsymbol{\theta}_l) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \cdots d\boldsymbol{\theta}_t. \tag{9.98}$$

This optimal choice is difficult to utilize in practice as it involves knowledge of the ability to draw from each of the distributions in the sequence.

This choice of optimal backward kernel is easily understood by interpreting it as the choice of kernel in which one would perform importance sampling on the space E rather than the product space E^t. The resulting incremental importance sampling weight for the optimal choice of backward kernel is simply

$$w_t(\boldsymbol{\theta}_{1:t}) = \frac{\pi_t(\boldsymbol{\theta}_t)}{\nu_t(\boldsymbol{\theta}_t)}. \tag{9.99}$$

In addition, we also note some examples of possible choices of the backward kernel given below along with the corresponding incremental importance sampling weights.

1. **Mixture Backward Kernel.** Given a mixture mutation kernel in Equation 9.94, the equivalent backward kernel is given by

$$L_{t-1,m}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \sum_{m=1}^{M} \beta_{t-1,m}(\boldsymbol{\theta}_t) L_{t-1,m}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) \tag{9.100}$$

with $\beta_{t,m}(\boldsymbol{\theta}_t) > 0$ and $\sum_{m=1}^{M} \beta_{t,m}(\boldsymbol{\theta}_t) = 1$. In this case, the incremental importance sampling weight can be written in the following form, with respect to an index auxiliary random variable for the mixture I_t that was sampled,

$$w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t, i_t) = \frac{\pi_t(\boldsymbol{\theta}_t)\beta_{t-1,i_t}(\boldsymbol{\theta}_t) L_{t-1,i_t}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1})\alpha_{t,i_t}(\boldsymbol{\theta}_{t-1}) M_{t,i_t}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}. \tag{9.101}$$

2. **Approximate Optimal Backward Kernel.** One of the best possible approximations to the optimal backward kernel is to consider replacing $\nu_t(\boldsymbol{\theta}_t)$ with $\pi_t(\boldsymbol{\theta}_t)$, to get

$$L_{t-1}^{opt}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \frac{\pi_{t-1}(\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{\int \pi_{t-1}(d\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}; \tag{9.102}$$

this would give an incremental importance weight of

$$w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \frac{\pi_{t-1}(\boldsymbol{\theta}_{t-1})}{\int \pi_{t-1}(\boldsymbol{\theta}_{t-1}) M_{t,i_t}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) d\boldsymbol{\theta}_{t-1}}. \tag{9.103}$$

Note that if resampling has occurred at time $t-1$, then this kernel is already equivalent to the optimal choice, and therefore, its particle approximation is already the optimal option. In general, if using the optimal backward kernel, one would still need to typically be able to approximate the univariate integrals, usually done via approximation using the particles at time $t-1$ as follows:

$$\int \pi_{t-1}(\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) d\boldsymbol{\theta}_{t-1} \approx \sum_{i=1}^{N} W_{t-1}^{(i)} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t). \tag{9.104}$$

Note that this results in an $O(N^2)$ algorithm, which is not ideal computationally.

3. **MCMC Backward Kernel.** A generic approximation of the 'approximate optimal backward kernel' in Equation 9.102 is often selected as an MCMC kernel in which one uses for the mutation kernel M_t an invariant MCMC kernel for target distribution π_t and the backward kernel given by

$$L_{t-1}(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \frac{\pi_t(\boldsymbol{\theta}_{t-1}) M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{\pi_t(\boldsymbol{\theta}_t)}. \tag{9.105}$$

This choice is a good approximation whenever the sequence of distributions π_{t-1} and π_t is close for all t, because this choice simply correspond to the time-reversed Markov kernel of the mutation kernel M_t. In addition, we note that you cannot adopt this choice for examples such as the successive sequence of constrained distributions as in the rare-event setting. When this backward kernel is utilized, one obtains an incremental importance weight given by a very simple form

$$w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \frac{\pi_{t-1}(\boldsymbol{\theta}_{t-1})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1})}. \tag{9.106}$$

9.7.2 INCORPORATING PARTIAL REJECTION CONTROL INTO SMC SAMPLERS

It is well known that the performance of SMC methods is strongly dependent on the mutation kernel. If M_t is poorly chosen, such that it does not place particles in regions of the support of π_t with high density, then many importance sampling weights will be close to zero. This leads to sample degeneracy, as a few well-located particles with large weights dominate the particle population, resulting in large variance for estimates made using these samples.

In Liu (2008) and Liu *et al.* (1998), a method is introduced, known as partial rejection control strategy, to overcome particle degeneracy in a sequential importance sampling setting. Under this mechanism, when the weight of a particle at distribution π_t falls below a finite threshold, $c_t \geq 0$, the particle is probabilistically discarded. It is replaced with a particle drawn from the previous distribution π_{t-1} which is then mutated to π_t. This new particle's weight is then compared to the threshold, with this process repeating until a particle is accepted. This concept was extended into an understanding of the resulting mutation kernel and developed under an SMC sampler framework in Peters *et al.* (2012). This approach is termed *partial* rejection, as the replacement particle is drawn from π_{t-1}, not π_1, see Liu (2008).

As demonstrated in Peters *et al.* (2012), under the SMC sampler framework, one may modify this approach and incorporate the partial rejection mechanism directly within the mutation kernel. Hence, at time $t-1$, the particle $\boldsymbol{\theta}_{t-1}$ is moved via the mutation kernel $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ and weighted according to Equation 9.91. This particle is accepted with probability p, determined by the particle's weight and the weight threshold c_t. If rejected, a new particle is obtained via the mutation kernel M_t, until a particle is accepted.

For the sequence of distributions π_t, $t = 1, \ldots, T$, the mutation and backward kernels M_t and L_{t-1}, a sequence of weight thresholds c_t, and PRC normalizing constants $r(c_t, \boldsymbol{\theta}_{t-1})$ (defined later), the SMC sampler PRC algorithm is given in Algorithm 9.4.

Algorithm 9.4 (SMC Sampler PRC Algorithm)

1. *Initialization:*
   Set $t = 1$.
   For $i = 1, \ldots, N$, *sample* $\boldsymbol{\theta}_1^{(i)} \sim \pi_1(\boldsymbol{\theta})$, *and set weights* $W_1^{(i)} = \frac{1}{N}$.
2. *Resample:*
   Normalize the weights $\sum_i W_t^{(i)} = 1$. *If* $[\sum_i (W_t^{(i)})^2]^{-1} < H$, *resample* N *particles with respect to* $W_t^{(i)}$ *and set* $W_t^{(i)} = \frac{1}{N}, i = 1, \ldots, N$.
3. *Mutation and correction:*
   Set $t = t + 1$ *and* $i = 1$*:*

(a) Sample $\boldsymbol{\theta}_t^{(i)} \sim M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t)$*, and set weight for* $\boldsymbol{\theta}_t^{(i)}$ *to*

$$W_t^{(i)} = W_{t-1}^{(i)} \frac{\pi_t(\boldsymbol{\theta}_t^{(i)}) L_{t-1}(\boldsymbol{\theta}_t^{(i)}, \boldsymbol{\theta}_{t-1}^{(i)})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1}^{(i)}) M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t^{(i)})}.$$

(b) With probability $1 - p^{(i)} = 1 - \min\{1, W_t^{(i)}/c_t\}$*, reject* $\boldsymbol{\theta}_t^{(i)}$ *and go to (a).*

(c) Otherwise, accept $\boldsymbol{\theta}_t^{(i)}$ *and set* $W_t^{(i)} = W_t^{(i)} r(c_t, \boldsymbol{\theta}_{t-1}^{(i)})/p^{(i)}$.

(d) Increment $i = i + 1$*. If* $i \leq N$*, go to (a).*

(e) If $t < T$*, go to Resample.*

Remark 9.12 *In the above SMC sampler PRC algorithm, we present the general framework in which we consider adaptive resampling. The derivation of the resulting normalizing constant for the PRC mechanism can be addressed under both adaptive and non-adaptive resampling schemes can be found in Peters* et al. *(2009, section 2.3). However, as they discuss, it will be shown to be computationally convenient when estimating the normalizing constant under PRC to consider the special case of H=N, thereby resampling at each iteration t.*

The above algorithm without the mutation and correction steps (b) and (c) is equivalent to the standard SMC sampler algorithm. In the resample stage, the degeneracy of the particle approximation is quantified through the usual estimate of the effective sample size, $1 \leq [\sum_i (W_t^{(i)})^2]^{-1} \leq N$ see Liu & Chen (1998). The addition of a rejection step at each time t effectively modifies the mutation kernel M_t. We denote the resulting kernel by M_t^*, to the choice presented in Equation 9.95. Thus the SMC sampler PRC algorithm can be considered as an SMC sampler algorithm with the mutation kernel $M_t^*(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ and the correction weight

$$W_t = W_{t-1} \frac{\pi_t(\boldsymbol{\theta}_t) L_{t-1}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1})}{\pi_{t-1}(\boldsymbol{\theta}_{t-1}) M_t^*(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}. \tag{9.107}$$

Remark 9.13

- ***Estimation of the Normalizing Constant.*** *As the normalizing constant* $r(c_t, \boldsymbol{\theta}_{t-1})$ *in the weight calculation (Equation 9.107) in general depends on* $\boldsymbol{\theta}_{t-1}$*, it must be evaluated as it will not disappear in the renormalization across all weights for each particle. Where no analytic solution can be found, approximating Equation 9.96 may be achieved by, for example, quadrature methods if the sample space* E *is relatively low dimensional or Monte Carlo methods if* E *is high dimensional.*
- ***Exact Kernel Selection Normalization.*** *It is an alternative approach which restricts the mutation and backward kernel choices and admits an exact solution for the normalizing constant. Furthermore, this approach provides a computationally efficient approach to dealing with the PRC normalizing constant. This involves selecting kernels* M_t *and* L_{t-1} *such that* $r(c_t, \boldsymbol{\theta}_{t-1}) = r(c_t)$ *will be constant for all particles* $\boldsymbol{\theta}_{t-1}$*. In this case, the value of* $r(c_t)$ *may be absorbed into the proportionality constant of the weights and safely ignored. Equation 9.96 suggests that this can be achieved if* $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$*,* W_{t-1} *and* $w(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)$ *are independent of* $\boldsymbol{\theta}_{t-1}$*.*

Specifying mutation kernels M_t such that $M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = M_t(\boldsymbol{\theta}_t)$ amounts to choosing a *global* kernel which is the same for all particles $\boldsymbol{\theta}_{t-1}$. The particle-dependent weight W_{t-1}

can be set to $1/N$ for all particles following a resampling (or pre-selection) step; hence, setting $H = N$ will induce resampling at each iteration. Finally, consider for a moment, the backward kernel of the form

$$L_{t-1}^{opt}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \frac{\pi_{t-1}(\boldsymbol{\theta}_{t-1})M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{\int \pi_{t-1}(\boldsymbol{\theta}_{t-1})M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)d\boldsymbol{\theta}_{t-1}}. \tag{9.108}$$

Under the backward kernel (Equation 9.108), the incremental weight can be approximated by

$$\begin{aligned} w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) &= \pi_t(\boldsymbol{\theta}_t) / \int \pi_{t-1}(\boldsymbol{\theta}_{t-1})M_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)d\boldsymbol{\theta}_{t-1} \\ &\approx \pi_t(\boldsymbol{\theta}_t) / \sum_{i=1}^{N} W_{t-1}^{(i)} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t). \end{aligned} \tag{9.109}$$

Under a global mutation kernel $M_t(\boldsymbol{\theta}_t)$, and following a resampling step, the incremental weight under this backward kernel reduces to $w_t(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t) = \pi_t(\boldsymbol{\theta}_t)/M_t(\boldsymbol{\theta}_t)$, which is independent of $\boldsymbol{\theta}_{t-1}$.

Remark 9.14 *One such example of a global mutation kernel one may consider involves* $M_t(\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1:N}, \boldsymbol{\theta}_t) = \sum_{i=1}^{N} W_{t-1}^{(i)} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t)$. *Thus, the weight calculation in Equation 9.107 becomes*

$$\begin{aligned} W_t &\propto \pi_t(\boldsymbol{\theta}_t) / \left[\min\left\{1, \frac{w(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{Nc_t}\right\} M_t(\boldsymbol{\theta}_t) \right] \\ &= \begin{cases} \pi_t(\boldsymbol{\theta}_t)/M_t(\boldsymbol{\theta}_t), & \text{if } \min\left\{1, \frac{w(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t)}{Nc_t}\right\} = 1, \\ Nc_t, & \text{otherwise.} \end{cases} \end{aligned}$$

It is instructive to consider the implications of this finding; firstly, the resulting acceptance probability for each particle will range over the interval $(0, 1)$. To see this, consider two illustrative scenarios, the first involves the trivial case of simply setting the user-controlled threshold to $c_t = 1/N$, thereby ensuring that as N increases, the acceptance probability does not necessarily decrease. This may not always be desirable because it reduces the threshold condition that particles must satisfy for large particle systems. The second non-trivial setting is to consider the incremental weight expression obtained in Equation 9.109. Under these choices for mutation and backward kernel, and assuming resampling in the setting $H = N$, we obtain an expression for the PRC probability of acceptance given by

$$\begin{aligned} \min\left\{1, \frac{w\left(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t\right)}{Nc_t}\right\} &= \min\left\{1, \frac{\pi_t(\boldsymbol{\theta}_t)}{Nc_t \sum_{i=1}^{N}(1/N)M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t)}\right\} \\ &= \min\left\{1, \frac{\pi_t(\boldsymbol{\theta}_t)}{c_t \sum_{i=1}^{N} M_t(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{\theta}_t)}\right\}. \end{aligned} \tag{9.110}$$

Note that under this setting, the SMC sampler PRC algorithm can be considered as a sequence of importance sampling strategies with partial rejection control.

Finally, we observe that there are several variants of the SMC samplers algorithm available in the context of interacting SMC samplers, annealed importance sampling and population Monte Carlo, island models and trans-dimensional SMC samplers, see examples in Jasra *et al.* (2007), Jasra *et al.* (2008), Neal (2001) and Cappé *et al.* (2004).

9.7.2.1 Finite Sample (Non-Asymptotic) Accuracy for Particle Integration. In this section, we detail some properties of the class of SMC algorithms discussed earlier, in particular what is known about the accuracy of such methods. In addition, we also present examples for estimators of quantiles of annual loss distributions from such approaches, of direct interest to capital estimation. We begin by presenting some recently presented examples of concentration inequalities for particle methods which are finite sample results, see discussion and references in Del Moral *et al.* (2013).

The exponential concentration inequalities presented below are satisfied under some regularity conditions on the particle weights and the mutation kernel M_n when defined on some general state space E_n, see specific probabilistic details of these conditions in Del Moral (2004).

Using the concentration analysis of mean field particle models, the following exponential estimate can be obtained, see discussion in Del Moral (2004) and references therein.

Theorem 9.17 (Finite Sample Exponential Concentration Inequality) *For any $x \geq 0$, $t \geq 0$, and any population size $N \geq 1$, the probability of the event*

$$\mathbb{P}\mathrm{r}\left[\left|\pi_t^N(\varphi) - \pi_t(\varphi)\right| \leq \frac{c_1}{N} \quad (1 + x + \sqrt{x}) + \frac{c_2}{\sqrt{N}} \quad \sqrt{x}\right] \geq 1 - e^{-x}, \tag{9.111}$$

where one defines

$$\pi_t^N(\varphi) = \sum_{i=1}^{N} W_t^{(i)} \varphi(\boldsymbol{\theta}_t^{(i)}), \quad \textit{and}$$

$$\pi_t(\varphi) = \int \varphi(\boldsymbol{\theta}_t)\pi_t(\boldsymbol{\theta}_t)d\boldsymbol{\theta}_t. \tag{9.112}$$

In the case of a stable SMC algorithm, that is, one that is insensitive to initial conditions, such as those we discuss earlier, the constants c and (c_1, c_2) do not depend on the time parameter. One can also bound the difference between the particle estimate of the target distribution and the true distribution as follows. Consider for any $\boldsymbol{\theta} = (\theta_i)_{1 \leq i \leq d}$ and any $(-\infty, x] = \prod_{i=1}^{d}(-\infty, \theta_i]$ cells in $E_t = \mathbb{R}^d$, we let

$$F_t(x) = \pi_t(\mathbb{I}_{(-\infty,x]}) \quad \text{and} \quad F_t^N(x) = \pi_t^N(\mathbb{I}_{(-\infty,x]}).$$

Using these definitions of the empirical particle constructed distribution function and the target distribution function at sequence number t in the sequence of distribution $\{\pi_1, \pi_2, \ldots, \pi_T\}$, then we can state the following corollary for the distribution functions for sequence of densities π_t given previously.

Corollary 9.2 *For any $y \geq 0$, $t \geq 0$, and any population size $N \geq 1$, the probability of the following event:*

$$\sqrt{N}\|F_t^N - F_t\| \leq c\sqrt{d(y+1)},$$

is greater than $1 - e^{-y}$.

This concentration inequality ensures that the particle repartition function F_t^N converges to F_t, almost surely for the uniform norm. We complete this section with an example of a non-asymptotic estimate for a risk measure estimation via SMC samplers output.

EXAMPLE 9.5 SMC Samplers Estimators for Risk Measures

Consider the single risk measure, where $d = 1$. Then let $F_t^{\leftarrow}$ be the generalized inverse on $[0, 1]$ of the function F_t which is the annual loss distribution for the LDA model under consideration; that is, we have that

$$F_t^{\leftarrow}(\alpha) := \inf\{\theta \in \mathbb{R} : F_t(x) \geq \alpha\}. \tag{9.113}$$

Now let $F_t^{\leftarrow}(\alpha) = q_{t,\alpha}$ be the quantile, of order α, and we denote by ζ_t^i the order particle statistic associated with the particle system θ_t^i at time t; that is, we have that

$$\zeta_t^1 := \theta_t^{\sigma(1)} \leq \zeta_t^2 := \theta_t^{\sigma(2)} \leq \ \ldots \ \leq \zeta_t^N := \theta_t^{\sigma(N)},$$

for some random permutation σ. We also denote by $q_{t,\alpha}^N := \zeta_t^{1+\lfloor N\alpha \rfloor}$ the α-particle quantile. By construction, we have that

$$\begin{aligned} |F_t(q_{t,\alpha}^N) - F_t(q_{t,\alpha})| &\leq |F_t(q_{t,\alpha}^N) - F_t^N(q_{t,\alpha}^N)| + |F_t^N(q_{t,\alpha}^N) - \alpha| \\ &\leq \|F_t^N - F_t\| + \left(\frac{1 + \lfloor N\alpha \rfloor}{N} - \alpha\right) \\ &\leq \|F_t^N - F_t\| + 1/N. \end{aligned} \tag{9.114}$$

This clearly implies that $q_{t,\alpha}^N$ converges almost surely to $q_{t,\alpha}$, as N tends to ∞. In addition, for any $y \geq 0$, $n \geq 0$, and any population size $N \geq 1$, the probability of the following event:

$$\sqrt{N}|F_n(q_{n,\alpha}^N) - \alpha| \leq c\sqrt{d(y+1)} + \frac{1}{\sqrt{N}},$$

is greater than $1 - e^{-y}$. ■

9.8 Multi-Level Sequential Monte Carlo Samplers for Higher Order Tail Expansions and Continuous Severity Distributions (Fixed *n*)

In this section, we discuss application of the SMC samplers methodology in the context of OpRisk, particular to evaluate the sequences of integrals present in the three approximation stages discussed in Section 9.6. In this section, we consider the case in which the number of

losses in a given year is fixed at n; hence, we are working with distributions comprised n-fold convolutions of the severity model. This could be the distribution, the density, the tail or functionals with regard to the n-fold convolution identity. In this setting, we aim to evaluate several important quantities for OpRisk via advanced Monte Carlo procedures which are efficient and exploit properties of previously developed recursive relationships. The first of these is to obtain samples from the distribution of the partial sum $Z_n = X_1 + \ldots + X_n$, which include a reasonable estimate of the tail functionals. In addition, we wish to quantify the accuracy and study the rate of convergence of tail approximation terms, in the series expansion of the n-fold convolution as discussed earlier; therefore, we will require samples not only from $f^{(n)*}(x)$ but also from a sequence of distributions given by $\{f^{(k)*}(x)\}_{1\leq k\leq n}$ in order to solve for each of the integrals in the decomposition.

Therefore, we see that we require a means to sample from such a sequence of distributions efficiently, including obtaining samples from the tails of each distribution in the sequence. Fortunately, as just discussed in the previous section, there exists a means of performing such a simulation via SMC samplers that will be discussed in detail later for this application. Then, after describing the SMC sampler framework, we provide the key components specific to the problem estimations in stages 1 to 3 in Section 9.5 that must be considered when applying or designing such an SMC sampler strategy. To achieve this, we demonstrate strategies for developing SMC-sampling-based estimators to represent the distribution F_{Z_n} and its tails via a set of weighted samples. These are specific extensions of the components presented for the generic SMC frameworks discussed in the estimation chapter in Cruz *et al.* (2015).

In particular, we have chosen to denote the application of SMC samplers in the context of OpRisk modeling examples as multi-level because it will typically require sequences of distributions on two or more algorithmic time scales:

- The first 'lowest level' scale will be for a given loss process, one may wish to obtain samples from a sequence of densities $\{\pi_t\}_{t=1}^{T_1}$ given, for example, by the single-loss process for risk cell i, given n-losses in the year according to $Z_n^{(i)} = \sum_{j=1}^n X_j^{(i)}$, which is characterized by marginal distribution sequence

$$\pi_t^{(i)}(x; n, q_t) = f_{Z_n^{(i)}}\left(z|\sum_{i=1}^n X_i > q_t\right) = *_{j=1}^n f_X^{(i)}(z)\mathbb{I}\left[\sum_{i=1}^n X_i > q_t\right], \quad (9.115)$$

  for sequence $0 < q_1 < q_2 < \ldots < q_T$.
- An example of a second 'intermediate level' scale could then be that this needs to be performed for an increasing number of losses such as $n = 1, 2, 3, \ldots$ producing a sequence of distributions for each n and for each threshold q_t given by

$$\{\pi_t^{(i)}(x; 1, q_1), \pi_t^{(i)}(x; 1, q_2), \ldots, \pi_t^{(i)}(x; 1, q_T), \pi_t^{(i)}(x; 2, q_1), \ldots, \pi_t^{(i)}(x; n, q_T)\}.$$

- An example of a third 'top level' scale could then be given by the fact that one wants to produce samples from a joint distribution characterized by the sequence $\{\pi_t(x; d, n, q_t)\}$ where $\pi_t(x; D, n, q_t)$ is given for aggregate annual loss random variable $Z_D = \sum_{j=1}^d \sum_{i=1}^n X_i^{(j)}\mathbb{I}\left[\sum_{i=1}^n X_i^{(j)} > q_t\right]$ assuming, for instance, independent loss

processes by distribution

$$\pi_t(x;d,n,q_t) = *_{j=1}^{d} *_{i=1}^{n} \pi_t(x;i,j,q_t) = *_{j=1}^{d} f_{X^{(j)}}^{(n)*}(x)\mathbb{I}\left[\sum_{i=1}^{n} X_i^{(j)} > q_t\right]. \tag{9.116}$$

9.8.1 KEY COMPONENTS OF MULTI-LEVEL SMC SAMPLERS

As briefly mentioned in the section introducing SMC samplers, one of the major difficulties with SMC-type algorithms is particle depletion, in which the weights of the majority of the particles gradually decrease to zero, while a few particle weights dominate the population. This can severely increase the variability of Monte Carlo estimates of expectations under each distribution in the sequence π_k. This depletion will be particular problematic when the sequence of distributions requires samples at each stage from the tails of the distribution for evaluation of risk measures. This issue and the design of a suitable sampling proposal and mechanism has been the focus of attention in the risk and insurance literature as well as the rare-event literature.

In this section, we discuss two approaches to the problem of designing such algorithms for sampling from a sequence of distributions in which an emphasis is placed on obtaining at each stage of the sequence, samples deep in the tails of the distribution. The first approach involves modification of the particle proposal distribution (known generically as the mutation kernel in the SMC literature) and the second approach involves a multi-level approach based on modification of the sequence of target distributions at each stage of the sequence.

The primary choices considered here for the sequence of distributions $\{\pi_k(x)\}_{1\leq k\leq n}$ are based on the discussions presented earlier and relate to the approximations described previously. Here four different distributional representations are considered to construct the sequence of distributions recursively (up to normalization):

1. **Representation 1** which is exact and considers a sequence of distributions given by

$$\pi_k^{(1)}(x) = f^{(k)*}(x) = \int_0^{\infty} f^{(k-1)*}(x-\tau)f(\tau)d\tau,\ \forall k \in \{1,2,\ldots,n\}.$$

2. **Representation 2** which is also exact and was derived from differentiation of the tail decomposition of the n-fold distribution function and given by

$$\pi_k^{(2)}(x) = f^{(k)*}(x) = T_{F,w}f^{(k-1)*}(x) + T_{F^{(k-1)*},1-w}f(x),\ \forall k \in \{1,2,\ldots,n\}.$$

3. **Representation 3** which is approximate (only asymptotically in $x \to \infty$ exact) and was derived from differentiation of the tail decomposition of the n-fold distribution function and given by

$$\pi_k^{(3)}(x) = f^{(k)*}(x) = T_{F,w}f^{(k-1)*}(x) + T_{\widetilde{F},1-w}f(x),\ \forall k \in \{1,2,\ldots,n\},$$

where $\widetilde{F}(x) = F(x)(1-k) - k$.

4. **Representation 4** the sequence given by Representation 4 is approximate (only asymptotically in $x \to \infty$ exact) and given by

$$\pi_k^{(4)}(x) = f^{(k)*}(x) = \frac{k}{k-1} T_{F,1/2} f^{(k-1)*}(x) + O\left(\overline{F}\left(\frac{x}{2}\right)^{1-\varepsilon} f\left(\frac{x}{2}\right)\right),$$
$$\forall k \in \{1, 2, \ldots, n\}.$$

9.8.1.1 Evaluating Target Distributions π_k Pointwise (up to Normalization). The SMC sampler algorithm proposed will require the evaluation of the target distributions π_k at each step in the sequence of distributions pointwise up to an unknown normalizing constant. This is clearly a challenging task in the context of the sequence of distributions developed for choices of distribution sequences defined by $\pi_k^{(1)}$ to $\pi_k^{(4)}$. Therefore, it is becoming popular in the SMC literature to address such evaluation challenges under one of three methods: the first involves a framework aimed at completely circumventing any direct evaluation of the target distributions, known as approximate Bayesian computation (ABC) (see discussions in the context of SMC samplers in Peters *et al.* (2012), Del Moral *et al.* (2012) and Jasra *et al.* (2012)); the second involves utilizing the previous set of particle estimates for the $(k-1)$th distribution in the sequence to aid in the evaluation of the target distribution in the kth iteration, this will work well in certain model settings; and the third involves numerical approximation of the target distribution at each stage. In this chapter, we consider the second and third methods, as it is not the intention to develop the ABC direction here. Instead we consider two choices for the approximate evaluation of the target distribution.

9.8.1.2 Approximation of the Target Distribution Method 1. The first is given by utilizing a previous particle approximation (set of N weighted Dirac measures) given for the p-th choice of distribution sequence by

$$\pi_{k-1}^{(p),N}(x) = \sum_{i=1}^{N} W_{k-1}^{(i)} \delta_{x_{k-1}^{(i)}}(x_{k-1}), \ \forall p \in \{1, 2, 3, 4\}. \tag{9.117}$$

Substitution of this approximation into the recursive expression for the sequence of distributions considered results in the following four approximate target distributions at the k-th iteration of the sequence:

$$\begin{aligned}
\widehat{\pi}_k^{(1)}(x) &= \int_0^{\infty} \pi_{k-1}^{(1),N}(x-\tau) f(\tau) d\tau = \sum_{i=1}^{N} W_{k-1}^{(i)} f(x - x^{(i)}), \\
\widehat{\pi}_k^{(2)}(x) &= T_{F,w} \pi_{k-1}^{(2),N}(x) + T_{\pi_{k-1}^{(2),N}(x),1-w} f(x) \\
&= \int_0^{wx} \pi_{k-1}^{(2),N}(x-\tau) f(\tau) d\tau + \int_0^{(1-w)x} f(x-\tau) \pi_{k-1}^{(2),N}(\tau) d\tau \\
&= \sum_{i=1}^{N} W_{k-1}^{(i)} f(x - x^{(i)}) \mathbb{I}[x_k^{(i)} \geq x(1-w)] \\
&\quad + \sum_{i=1}^{N} W_{k-1}^{(i)} f(x - x^{(i)}) \mathbb{I}[x_k^{(i)} \leq (1-w)x],
\end{aligned}$$

$$\begin{aligned}
\widehat{\pi}_k^{(3)}(x) &= T_{F,w}\pi_{k-1}^{(3),N}(x) + T_{\widetilde{F},1-w}f(x) \\
&= \int_0^{wx} \pi_{k-1}^{(3),N}(x-\tau)f(\tau)d\tau + \int_0^{(1-w)x} f(x-\tau)d\widetilde{F}(\tau) \\
&= \sum_{i=1}^{N} W_{k-1}^{(i)} f(x-x^{(i)})\mathbb{I}[x_k^{(i)} \geq x(1-w)] + (1-k)\int_0^{(1-w)x} f(x-\tau)f(\tau)d\tau, \\
\widehat{\pi}_k^{(4)}(x) &= \frac{k}{k-1}T_{F,1/2}\pi_{k-1}^{(4),N}(x) \\
&= \frac{k}{k-1}\int_0^{x/2} \pi_{k-1}^{(4),N}(x-\tau)f(\tau)d\tau \\
&= \frac{k}{k-1}\sum_{i=1}^{N} W_{k-1}^{(i)} f(x-x^{(i)})\mathbb{I}[x_k^{(i)} \geq x/2].
\end{aligned}$$

It is clear from these representations that it is now possible to evaluate pointwise the target distribution and, therefore, the importance sampling weights at each stage of the SMC samplers algorithm. It should also be noted that the choice of sequence of distributions given in option $\widehat{\pi}_k^{(3)}(x)$ also requires the exact or approximate evaluation of the one-dimensional integral once offline to obtain the function $(1-k)\int_0^{(1-w)x} f(x-\tau)f(\tau)d\tau$; this can be done exactly or via quadrature.

9.8.1.3 Approximation of the Target Distribution Method 2. The second approach involves discretizing the severity distribution and utilizing the recursions presented earlier to recursively evaluate the n-fold convolutions. Then given the probability mass functions over a discrete grid the approximations given under the 'Approximation of the Target Distribution Method 1' can be replaced with the deterministic grid counterparts.

Remark 9.15 *Under either of the above two methods, we have obtained in the first instance a randomized particle grid importance weighted approximation to solve for the required evaluations of the sequence of densities in the SMC sampler. In the second case, the target distributions are evaluated based on discretizations and the De Pril transforms to approximate the target sequence of distributions. Either of these approximations can then be utilized in the SMC samplers framework when evaluating the SMC sampler weights. It should be noted that in this case, the samples from the mutation kernel are still anywhere on the support (not refined to a deterministic grid) as would be the case in a standard discretization method.*

Hence, having recalled the basic method of SMC samplers and specified the choice of sequence of target distributions under four possibilities, we then explained two different approaches to how to evaluate the target distributions. This then allows one to proceed with developing any particular choice of SMC sampler where the choice of mutation and backward kernel can be explored as well as possibilities of resample move and PRC strategies, the interested reader is referred to the extensive detail in Cruz *et al.* (2015, chapter 7).

Next we move from considerations of partial sums for known numbers of losses to the case of random sums (compound processes) and the recursions available in these cases for both discretized severity distribution approaches as well as continuous recursion approaches that will again adopt specialized SMC and importance sampling solutions.

9.9 Recursions for Compound Process Distributions and Tails with Discretized Severity Distribution (Random N)

There are two common approaches adopted in the literature for performing recursive evaluation of compound processes, the approach of Panjer and the approximation approach of Kornya, see discussion on the relationship in Kornya (2007). In this section, we focus on the Panjer recursion as it is generally most widely used in practice.

The evaluation of distributions of random sums or compound processes has been ubiquitous in actuarial mathematics for many decades (see, for instance, Klugman *et al.* (1998), Panjer (1981) and Klugman *et al.* (1998). It typically arises as the preferred method of modeling the probability of ruin and the loss distributions for insurance claims that can be assumed to arise according to, for example, a Poisson or negative binomial compound process. In the OpRisk setting, the random sum models the total annual loss that is incurred in the given risk process, business unit structure.

Hence, such random sums are typically considered when modeling the distribution of the total claims incurred in a fixed period of time. What makes the explicit computation of these loss distributions difficult is that the conditional distribution of the amount of total claims or loss given a certain number of claims n has occurred involves an n-fold convolution of the severity distribution. Additionally, because the number of claims is also now treated as a random variable, the combined LDA distribution is a mixture of all possible numbers of losses and the resulting distribution for each of the possible partial sums of the loss amounts. In this section, several recursive approaches to the evaluation and sampling of such distributions are developed.

Before proceeding with these recursive approaches, it is relevant to consider under which conditions such infinite series comprised positive terms will converge. In this case, the series one can consider corresponds to the both the partial sums $\{S_n\}_{n\geq 0}$ with $S_n = X_1 + \cdots + X_n$, where X_n are i.i.d. loss random variables from discretized distribution F and the number of losses $N = n$, probabilities $\{p_n\}_{n\geq 0}$. The existence of the distribution of the compound process Z_n clearly depends on the convergence of the sequences. We present a few basic results regarding convergence of such sequences below that are alternatives to the Kolmogorov three series theorem presented in Chapter 3.

Theorem 9.18 (Convergence of Infinite Positive Series) *Consider the sequence of positive-valued elements $\{x_n\}_{n\geq l}$ for some integer $l \geq 0$. Then a sequence of such positive numbers that satisfies the following limit for finite c,*

$$\lim_{n\uparrow\infty} \frac{x_n}{x_{n-1}} = c,$$

has for $c < 1$ convergence to a finite value of the infinite sum $\sum_{n=l}^{\infty} x_n$. Alternatively, if one can show the existence of the finite constant c for the limit,

$$\lim_{n\uparrow\infty} n\left(1 - \frac{x_n}{x_{n-1}}\right) = c,$$

then for $c > 1$, one has convergence to a finite value of the infinite sum $\sum_{n=l}^{\infty} x_n$.

See a simple proofs of these results in Sundt & Vernic (2009, p. 25–26). The final result on convergence of infinite sequences is more specific to the types of recursive expressions to be considered below as detailed in Theorem 9.19.

Theorem 9.19 *Consider the sequence of positive-valued elements* $\{x_n\}_{n\geq l}$ *for some integer* $l \geq 0$ *that satisfy the recursive expression given by*

$$x_n = \left(a + \frac{b}{n}\right) x_{n-1}, \quad n \in \{l+1, l+2, \dots\}, \tag{9.118}$$

with $a > 0$ *and* $b \geq -(l+1)a$. *Then one has for* $a < 1$ *convergence to a finite value of the infinite sum* $\sum_{n=l}^{\infty} x_n$ *and divergence if* $a > 1$. *If* $a = 1$, *convergence is achieved for* $b < -1$ *and divergence otherwise.*

Having detailed these conditions under which infinite summations will have convergent evaluations, one may proceed with understanding of how to developed recursive methods to evaluate such compound process (infinite sums) pointwise. This naturally leads one to the well-known and widely used class of the Panjer recursions described next.

9.9.1 PANJER RECURSIONS FOR COMPOUND DISTRIBUTIONS WITH DISCRETIZED SEVERITY DISTRIBUTIONS

If the severity distributions F are discrete, then efficient, deterministic techniques based on the z-transform may be employed, but this approach does not generalize to continuous severity distributions. Alternatively, the Panjer recursion provides a recursive expression for the evaluation of the coefficients c_k of the probability-generating function $P(z)$. The Panjer recursion in its standard form is designed for discrete severity distributions; hence, it is common in practice for practitioners to discretely approximate a continuous distribution in order to utilize the Panjer recursion. This could involve, for example, rounding all loss amounts to the nearest multiple of a monetary unit δ, such as the nearest USD 1000.

Before presenting the basic Panjer recursion, consider the estimation of the compound process distribution function in which we do not impose any requirements on the frequency distribution,

$$F_{Z_n}(x) = \sum_{k=0}^{\infty} p_k F^{(k)*}(x), \tag{9.119}$$

where

$$F^{(k)*}(x) = \int_0^x F^{(k-1)*}(x-\tau) f(\tau) d\tau, \tag{9.120}$$

with

$$F^{(0)*}(x) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \tag{9.121}$$

Now under the discretized representation of the distribution over a grid defined by increments of amount δ, one can proceeds according to the following defined quantities

$$f(k) = \mathbb{P}\mathrm{r}[X_i = k\delta], \quad p_k = \mathbb{P}\mathrm{r}[N = k], \quad f_{Z_N}(k) = \mathbb{P}\mathrm{r}[Z_N = k\delta], \tag{9.122}$$

to estimate the terms in this convolution via the recursions for the initialization $f_0 = 0$ for integers $k = 0, 1, 2, \ldots$ according to

$$f_{Z_N}(n) = \sum_{k=1}^{n} p_k f^{(k)*}(n), \; n \geq 1,$$
$$f_{Z_N}(0) = \Pr[Z_N = 0] = \Pr[N = 0] = p_0, \tag{9.123}$$

where $f^{(k)*}(n)$ is given by

$$f^{(k)*}(n) = \sum_{i=0}^{n} f^{(k-1)*}(n-i) f(i), \tag{9.124}$$

with $f^{(0)*}(0) = 1$ and $f^{(0)*}(n) = 0$ if $n \geq 1$.

At this stage, we note the following properties of this recursive solution:

1. If at the origin the severity distribution satisfies the condition that $f(0) = \Pr[X = 0] = 0$, then this implies that $f^{(k)*} = 0$ for $k > n$ and the above summation is up to n only.
2. If at the origin the severity distribution satisfies the condition that $f(0) = \Pr[X = 0] > 0$, then this implies that $f^{(k)*} > 0$ for all n and k. As a consequence, the upper limit in Equation 9.123 should be replaced with infinity.
3. The computational complexity of evaluating this recursion which involves the calculation of $f_{Z_N}(0), f_{Z_N}(1), \ldots, f_{Z_N}(n)$ via Equation 9.7 is of order $O(n^3)$ operations.

It is well known that if one considers the class of frequency distributions for the random number of losses N which belongs to one of the Panjer classes, then one can significantly improve the recursive evaluation of Equation 9.123 by the Panjer recursion as detailed in Theorem 9.20, see Panjer (1981). The Panjer recursions reduces the computational complexity for the evaluation of $f_{Z_N}(0), f_{Z_N}(1), \ldots, f_{Z_N}(n)$ from $O(n^3)$ to $O(n^2)$. In addition, the strong stability of the Panjer recursion is studied in Panjer & Wang (1993). It can be shown that the accumulated rounding error of the Panjer recursions increases linearly with n with a gradient never exceeding 1.

Theorem 9.20 (Panjer Recursions for Compound Processes with Discretized Severity) *If the frequency distribution (probability mass) function p_n satisfies membership of the Panjer $(a, b, 0)$ class with*

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \forall n \geq 1, \quad \text{and } a, b \in \mathbb{R}, \tag{9.125}$$

then the compound distribution in Equation 9.123 satisfies the simplified recursion

$$f_{Z_N}(n) = \frac{1}{1 - af(0)} \sum_{j=1}^{n} \left(a + \frac{bj}{n}\right) f(j) f_{Z_N}(n-j), \; \forall n \geq 1,$$
$$f_{Z_N}(0) = \sum_{k=0}^{\infty} (f(0))^k p_k. \tag{9.126}$$

If the frequency distribution (probability mass) function p_n *satisfies membership of the Panjer* $(a, b, 1)$ *class with*

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \forall n \geq 2, \quad \text{and } a, b \in \mathbb{R}, \tag{9.127}$$

then the compound distribution in Equation 9.123 satisfies the simplified recursion

$$f_{Z_N}(n) = \frac{(p_1 - (a+b)p_0)f(n) + \sum_{j=1}^{n}\left(a + \frac{bj}{n}\right) f(j) f_{Z_N}(n-j)}{1 - af(0)}, \ \forall n \geq 1, \tag{9.128}$$

$$f_{Z_N}(0) = \sum_{k=0}^{\infty} (f(0))^k p_k.$$

If the frequency distribution (probability mass) function p_n *satisfies membership of the Panjer* (a, b, l) *class with*

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \forall n \geq l+1, \quad \text{and } a, b \in \mathbb{R}, \tag{9.129}$$

with initial values $p_0 = \cdots = p_{l-1} = 0$ *and the case in which* $f(0) = 0$*, then the compound distribution in Equation 9.7 satisfies the simplified recursion*

$$f_{Z_N}(n) = p_l f^{(l)*}(n) + \sum_{j=1}^{n}\left(a + \frac{bj}{n}\right) f(j) f_{Z_N}(n-j), \ \forall n \geq l, \tag{9.130}$$

$$f_{Z_N}(0) = \sum_{k=0}^{\infty} (f(0))^k p_k.$$

To understand where this recursion is obtained, it will be useful to discuss its derivation. In this regard, we will derive the recursion from the continuous severity setting, and it is then trivial to obtain the discrete recursion through a discretization of the severity distribution. The recursion relies upon two identities:

- n-fold convolution identity

$$\int_0^x f_X(y) f^{(n)*}(x-y)\, dy = f^{(n+1)*}(x), \quad n = 1, 2, 3, \ldots \tag{9.131}$$

- Conditional expectations identity

$$\int_0^x \frac{y f_X(y) f^{(n)*}(x-y)}{f^{(n+1)*}(x)\, dy} = \frac{x}{n+1}, \quad n = 1, 2, 3, \ldots \tag{9.132}$$

First, consider the definition of a convolution for severity density functions given by

$$f^{(n+1)*}(x) = \int_0^x f(\tau) f^{(n)*}(x-\tau) d\tau, \quad \forall n = 1, 2, 3, \ldots \tag{9.133}$$

Now consider the partial sum given by $S_{n+1} = X_1 + X_2 + \ldots + X_{n+1}$. Then the conditional distribution of X_1 for the partial sum $S_{n+1} = x$ is given by

$$\begin{aligned} f_{X_1}(\tau|X_1 + \cdots + X_{n+1} = x) &= \frac{f_{X_1}(\tau) f_{X_1+\cdots+X_{n+1}}(x|X_1 = \tau)}{f_{X_1+\cdots+X_{n+1}}(x)} \\ &= \frac{f_{X_1}(\tau) f_{X_1+\cdots+X_n}(x-\tau)}{f_{X_1+\cdots+X_{n+1}}(x)} \\ &= \frac{f_{X_1}(\tau) f_{X_1}^{(n)*}(x-\tau)}{f_{X_1}^{(n+1)*}(x)}. \end{aligned} \tag{9.134}$$

Now consider the average of the first loss given that the partial sum $S_{n+1} = x$,

$$\begin{aligned} \mathbb{E}[X_1|X_1 + \cdots + X_{n+1} = x] &= \int_0^x \tau f_{X_1}(\tau|X_1 + \cdots + X_{n+1} = x) d\tau \\ &= \int_0^x \tau \frac{f_{X_1}(\tau) f_{X_1}^{(n)*}(x-\tau)}{f_{X_1}^{(n+1)*}(x)} d\tau. \end{aligned} \tag{9.135}$$

Note that the losses in the process are assumed to be i.i.d. from the severity distribution F, and hence, the conditional mean of any element in the sum with $n+1$ elements will be given by

$$\begin{aligned} \mathbb{E}[X_1|X_1 + \cdots + X_{n+1} = x] &= \frac{1}{n+1} \sum_{i=1}^{n+1} \mathbb{E}[X_i|X_1 + \cdots + X_{n+1} = x] \\ &= \frac{1}{n+1} \mathbb{E}[X_1 + \cdots + X_{n+1}|X_1 + \cdots + X_{n+1} = x] \\ &= \frac{x}{n+1}, \end{aligned} \tag{9.136}$$

When this result is equated with the expression for the conditional mean in Equation 9.135 , one obtains after rearranging

$$\frac{1}{n+1} f_{X_1}^{(n+1)*}(x) = \frac{1}{x} \int_0^x \tau f_{X_1}(\tau) f_{X_1}^{(n)*}(x-\tau) d\tau. \tag{9.137}$$

Now consider the compound loss density and introduce the Panjer class of claim number distributions for the frequency distribution given by the recursive relationship,

$$\Pr[N = n] = p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \tag{9.138}$$

for some a and b to obtain,

$$\begin{aligned} f_{Z_N}(x) &= \sum_{n=1}^{\infty} p_n f^{(n)*}(x) \\ &= p_1 f(x) + \sum_{n=2}^{\infty} p_n f^{(n)*}(x) \end{aligned}$$

$$
\begin{aligned}
&= p_1 f(x) + \sum_{n=1}^{\infty} p_{n+1} f^{(n+1)*}(x) \\
&= p_1 f(x) + \sum_{n=1}^{\infty} \left(a + \frac{b}{n+1} \right) p_n f^{(n+1)*}(x) \\
&= p_1 f(x) + a \sum_{n=1}^{\infty} p_n \int_0^x f(\tau) f^{(n)*}(x-\tau) d\tau \\
&\quad + b \sum_{n=1}^{\infty} p_n \frac{1}{x} \int_0^x \tau f(\tau) f^{(n)*}(x-\tau) d\tau \\
&= p_1 f(x) + a \int_0^x f(\tau) \sum_{n=1}^{\infty} p_n f^{(n)*}(x-\tau) d\tau \\
&\quad + \frac{b}{x} \int_0^x \tau f(\tau) \sum_{n=1}^{\infty} p_n f^{(n)*}(x-\tau) d\tau \\
&= p_1 f(x) + \int_0^x \left(a + \frac{by}{x} \right) f(\tau) f_{Z_N}(x-\tau) d\tau. \qquad (9.139)
\end{aligned}
$$

The condition for $f_{Z_N}(0)$ in the Panjer recursion in Theorem 9.20 corresponds to the probability-generating function of the random variable for the number of losses N, evaluated at $f(0)$. Furthermore, the condition on the frequency distribution in Equation 9.125 was proven in Sundt & Jewell (1981) to be satisfied for Poisson, negative binomial and binomial distributions with parameters for $f_{Z_N}(0)$, a and b in each case given in Table 9.1.

According to Hess *et al.* (2001, theorem 3.4), one can show that the Panjer Class $(a, b, 1)$ distributions include

1. zero-truncated distributions of $(a, b, 0)$ with $p_0 = 0$ such as zero-truncated Poisson, zero-truncated negative binomial and zero-truncated binomial.
2. zero-modified distributions of $(a, b, 0)$ with $p_0 > 0$ which have a modified probability of $N = 0$. Typically viewed as a mixture of $(a, b, 0)$ distribution with a degenerate distribution concentrated on zero.
3. Logarithmic distribution Log(q) for $q \in (0, 1)$ with $p_n = \frac{1}{|\ln(1-q)|} \frac{q^n}{n}$ for all $n \geq 1$ is also a member of the $(a, b, 1)$ Panjer class with $a = q$ and $b = -q$.

TABLE 9.1 Panjer recursion (a, b) parameters and starting values $f_{Z_N}(0)$

Frequency Distribution	a	b	$f_{Z_N}(0)$
Poisson(λ)	0	λ	$\exp(\lambda(f_0 - 1))$
negative binomial(r, q)	$1 - q$	$(1 - q)(r - 1)$	$\left(1 + (1 - f_0) \dfrac{1 - q}{q}\right)^{-r}$
binomial(m, q)	$-\dfrac{q}{1 - q}$	$\dfrac{q(m + 1)}{(1 - q)}$	$(1 + q(f_0 - 1))^m$

Also the Panjer class (a, b, l) distributions include $l-1$ truncated distributions of $(a, b, 0)$ with $p_0 = \cdots = p_{l-1} = 0$ and the case in which $f_0 = 0$ such as $l-1$ truncated Poisson, $l-1$ truncated negative binomial and $l-1$ truncated binomial.

Note that all these different Panjer classes can be considered as sub-classes of the Sundt–Jewell classes of distributions, see detailed discussion in Sundt & Vernic (2009, chapter 2). Further details on the Panjer classes and higher order Panjer recursions can be found in Hess *et al.* (2001) and the book of Sundt & Vernic (2009). In addition, we note that the De Pril recursions presented previously for the evaluation of an n-fold convolution can be directly related to the Panjer recursion, see detailed discussions in Sundt & Vernic (2009).

Before proceeding, we consider some examples given in Table 9.2 which presents results of Panjer recursion calculations of the $Poisson(100) - LogNormal(0, 2)$ compound distributions using central difference discretization with the step $\delta = 1$. Of course, the accuracy of the result depends on the step size as shown by the results for the 0.999 quantile and expected shortfall versus δ, see Table 9.3. It is, however, important to note that the error of the result is due to discretization only and there is no truncation error (i.e. the severity is not truncated by some large value). Also, note that the 0.999 quantile estimate is more accurate than the 0.999 shortfall estimate (for given time δ).

TABLE 9.2 Example of Panjer recursion calculating the $Poisson(100) - LogNormal(0, 2)$ compound distributions using central difference discretization with the step $\delta = 1$

n	$f(n)$	$f_{Z_N}(n)$	$\sum_{i=0}^{n} f_{Z_N}(i)$
0	0.364455845	2.50419×10^{-28}	2.50419×10^{-28}
1	0.215872117	5.40586×10^{-27}	5.65628×10^{-27}
2	0.096248034	6.07589×10^{-26}	6.64152×10^{-26}
⋮	⋮	⋮	⋮
5,847	2.81060×10^{-9}	4.44337×10^{-7}	0.998999329
5,848	2.80907×10^{-9}	4.44061×10^{-7}	0.998999773
5,849	2.80755×10^{-9}	4.43785×10^{-7}	0.999000217

TABLE 9.3 Convergence of Panjer recursion estimates $\widehat{q}_{0.999}$ and $\widehat{\mathrm{ES}}_{0.999}$ of the 0.999 quantile and expected shortfall, respectively, for the $Poisson(100) - LogNormal(0, 2)$ compound distributions using central difference discretization versus the step size δ

δ	N	$\widehat{q}_{0.999}$	$\widehat{\mathrm{ES}}_{0.999}$
2	2,921	5,842	20,131
1	5,849	5,849	13,519
0.5	11,703	5,851.5	10,831
0.25	23,411	5,852.75	9,873
0.125	46,824	5,853	9,575
0.0625	93,649	5,853.0625	9,494

Here, $N = \widehat{q}_{0.999}/\delta$ is the number of steps required.

Next we also consider the convergence properties numerically of the Panjer recursion.

These results show that if frequency is large, then underflow may occur in the computations. Underflow is the case when the numerical calculations produce a number outside the range of representable numbers leading to 0. It is easy to see in the case of $Poisson(\lambda)$ and $f_0 = 0$ when $f_{Z_N}(0) = \exp(-\lambda)$. In this case, the underflow will occur for $\lambda \gtrsim 700$ on a 32-bit computer with double precision calculations. Re-scaling $f_{Z_N}(0)$ by large factor γ to calculate the recursion (and de-scaling the result) will not resolve the issue because overflow will occur for $\gamma f_{Z_N}(n)$ (i.e. calculations will produce a number outside of representative range leading to ∞). The following identity helps to overcome this problem in the case of Poisson frequency:

$$F_{Z_N}^{(m)*}(z; \lambda/m) = F_{Z_N}(z; \lambda). \tag{9.140}$$

That is, calculate the compound distribution $F_{Z_N}(z; \lambda/m)$ for some large m to avoid underflow. Then perform m convolutions for the obtained distribution directly or via FFT; see Panjer & Willmot (1986). Similar identity is available for $NegBinomial(r, p)$:

$$F_{Z_N}^{(m)*}(z; r/m) = F_{Z_N}(z; r). \tag{9.141}$$

In the case of $Binomial(M, p)$,

$$F_{Z_N}^{(m)*}(z; m_1) * F_{Z_N}(z; m_2) = F_{Z_N}(z; M), \tag{9.142}$$

where $m_1 = \lfloor M/m \rfloor$ and $m_2 = M - m_1 m$. For numerical efficiency, one can choose $m = 2^k$ so that instead of m convolutions of $F_{Z_N}(\cdot)$ only k convolutions are required $F_{Z_N}^{(2)*}, F_{Z_N}^{(4)*}, \ldots, F_{Z_N}^{(2^k)*}$, where each term is the convolution of the previous one with itself.

If one now considers applications of the Panjer recursion, it is natural to question the behaviour of such recursive algorithms with regard to round-off errors in the representation of the solutions given a finite number of digits. It is natural to then ask questions relating to the stability of these recursive calculations using the Panjer recursion. In other words, asking 'How do errors propagate when using these recursive formulas?'

Initially, a study was performed on absolute errors of such a recursion by De Vylder & Goovaerts (1994) and Goovaerts & De Vylder (1984), and they concluded that under a criterion considering absolute error propagation such a recursive scheme is numerically unstable because of inflation of absolute errors. Then in Panjer & Wang (1993), they studied the propagated relative error compared to the solution and argued this is a more meaningful measure of stability of such recursive relationships. Under a relative error analysis, the Panjer recursions are in fact seen to be generally well behaved and numerically stable for most settings.

To address such questions, one considers if the terms on the right-hand side of the recursive expression are strictly positive, in which case it can be shown that errors will not propagate faster than linearly. However, if there are negatives present (or alternating signs) as will occur with the Binomial distribution, then errors can blow up, see discussions in Panjer & Wang (1993). In this paper, the authors use theory on recurrence equations and a relative error analysis to study concepts and criteria for the stability of a recurrence equation of the form discussed, see detailed discussion and examples in Panjer & Wang (1993).

From a practical perspective, although not of primary interest in the rare-event high consequence setting discussed in this chapter for heavy-tailed severity models, it can be possible in OpRisk models to have high frequency loss processes. In such settings, one can expect two

potential problems to arise. Firstly, there can be underflow when computing $f_{Z_N}(0)$ and the solution could be incorrectly set to zero. The other source of computational difficulty could be that aggregate distribution may in high frequency settings frequently use significantly more computer memory than the discretized severity distribution, potentially exceeding the size of the largest available storage.

Practically, one can for high frequency settings consider the fact that there is very little aggregate probability near zero. Hence, it may be acceptable to start the recursions at a reasonably safe value, for example, it is suggested in the literature to consider in practice 6 standard deviations below the mean. Then normalize the solution post processing to make the probabilities sum to 1.

So far, the introduction of the Panjer recursion has been for cases in which the Panjer class of counting distributions considered was of order (a, b, k) for $k = 0$. In general, one can generalize these results for a generalized Panjer recursion given in Theorem 9.21, see details in Hess *et al.* (2001, theorem 4.1)

Theorem 9.21 (Higher Order Panjer Class Recursions) *Given a non-degenerate frequency distribution $N \sim F_N(n)$, then for $a, b \in \mathbb{R}$ and $k \in \mathbb{N}_0$, one has equivalence in the following statements:*

- *The frequency distribution $F_N(n)$ is in the Panjer$(a, b; k)$ class.*
- *For every discretized frequency distribution $F_X = \{\widehat{f}_n\}_{n\in\mathbb{N}_0}$ with $\widehat{f}_0 = 0$ and for every $l \geq 1$, the probability-generating function for the annual loss $M_Z(z)$ for the compound annual loss distribution satisfies the differential equation defined with respect to the p.g.f. of the severity distribution $M_X(z)$ according to*

$$(1 - am_X(z))\frac{d^l}{dz^l}M_Z(z) = \sum_{i=1}^{l} \frac{l!}{(l-i)!i!}\left(a + b\frac{i}{l}\right)\frac{d^{l-i}}{dz^{l-i}}M_Z(z)\frac{d^i}{dz^i}M_X(z) + \mathbb{P}\mathrm{r}(N = k)\frac{d^l}{dz^l}M\sum_{i=1}^{k} X_i(z) \tag{9.143}$$

with $z \in [0, 1)$ and initial condition $\frac{d^j}{dz^j}M_Z(0) = 0$ for all $j \leq k - 1$.

9.9.2 ALTERNATIVES TO PANJER RECURSIONS: RECURSIONS FOR COMPOUND DISTRIBUTIONS WITH DISCRETIZED SEVERITY DISTRIBUTIONS

In some settings, it may be advantageous both from a computational efficiency as well as numerical accuracy or stability to consider alternative recursions or higher order recursions for compound processes. There are numerous other recursions available for compound distribution evaluation, some of the more useful variants are given below under assumptions on the frequency, the severity distribution or both.

An example of this is the recursion for a compound Poisson distribution with severity distribution satisfying that $f(0) = 0$, then one obtains a recursion given in Theorem 9.22.

Theorem 9.22 (Recursions for Compound Poisson Distributions with Discretized Severity) *Consider a compound Poisson distribution for a single risk LDA model in which the severity distribution $f(x)$ is discretized (w.l.o.g.) over the non-negative integers and satisfies $f(0) = 0$, then the compound distribution given by*

$$f_{Z_N}(x) = \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} \exp(-\lambda) f^{(n)*}(x) \tag{9.144}$$

is evaluated recursively according to

$$f_{Z_N}(x) = \frac{\lambda}{x} \sum_{y=1}^{x} y f(y) f_{Z_N}(x-y), \forall x \in \{1, 2, \dots\}, \tag{9.145}$$

with initialization

$$f_{Z_N}(0) = \exp(-\lambda). \tag{9.146}$$

Additionally, in Sundt & Vernic (2009, section 3.3), an alternative class of recursions for mixed Poisson compound distributions can be utilized if certain conditions are satisfied for the discretized severity distribution. If the severity distribution $f(x)$ after discretization satisfies that the first derivative of the power series representation of the probability mass function satisfies the relation

$$\frac{d}{ds}\mathbb{E}[s^X] = \int_0^{\infty} s^x \, dF(x) = \frac{\sum_{y=1}^{r} \eta(y) s^{y-1}}{1 - \sum_{y=1}^{r} \chi(y) s^y} \tag{9.147}$$

for some functions η and χ and r either a positive integer of infinity, then the recursion in Theorem 9.23 for the compound Poisson distribution evaluation holds.

Theorem 9.23 *If the frequency distribution is a Poisson distribution with rate λ and the severity distribution f is discretized and takes values with positive probability only on the non-negative integers (w.l.o.g.). Furthermore, assume that there exists function η and χ on $\{1, 2, \dots, r\}$ and an integer r either a positive integer of infinity that satisfies that the first derivative of the power series representation of the probability mass function satisfies the relation*

$$\frac{d}{ds}\mathbb{E}[s^X] = \int_0^{\infty} s^x \, dF(x) = \frac{\sum_{y=1}^{r} \eta(y) s^{y-1}}{1 - \sum_{y=1}^{r} \chi(y) s^y}, \tag{9.148}$$

then the compound distribution is evaluated according to the recursion

$$f_{Z_N}(x) = \sum_{y=1}^{r} \left(\frac{\lambda}{x} \eta(x) + \left(1 - \frac{y}{x}\right) \chi(y) \right) f_{Z_N}(x-y), \quad \forall x \in \{1, 2, \dots\}. \tag{9.149}$$

In addition, in cases in which the frequency distribution satisfies the recursive evaluation given by

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \forall n \in \{l, l+1, \dots, r\}, \tag{9.150}$$

then in this case, the compound distribution can be evaluated recursively according to

$$f_{Z_N}(x) = p_l f^{(l)*}(x) - \left(a + \frac{b}{r+1}\right) p_r f^{(r+1)*}(x) + \sum_{y=1}^{x} \left(a + b\frac{y}{x}\right) f(y) f_{Z_N}(x-y),$$
$$\forall x \in \{l, l+1, \dots\}. \tag{9.151}$$

To complete this section of recursions for evaluation of the compound process, we also mention the framework developed and known as Waldmann's recursion Waldmann (1996, theorem 1) which provides an alternative recursion one can consider. The advantage of the Waldmann recursion is that instead of working with the density of the compound process, it considers the distribution function which has the advantage that it is strictly monotone. The recursion proceeds according to Theorem 9.24.

Theorem 9.24 (Waldmann's Recursion for Compound Distributions with Discretized Severity) *Consider the compound process with discretized severity distribution given by $F_{Z_N}(x)$ for all $x \in \mathbb{N}_0$. If the frequency distribution satisfies the condition that*

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}, \quad \forall n \in \{1, 2, 3 \dots\}, \tag{9.152}$$

then the distribution of the compound process can be evaluated recursively according to

$$x F_{Z_N}(x) = r_1(x) + r_2(x), \tag{9.153}$$

where $F_{Z_N}(0) = p_0$ and for all $x \in \mathbb{N}_0$ one has

$$r_1(x) = r_1(x-1) + F_{Z_N}(x-1)$$
$$r_2(x) = a \sum_{i=1}^{x-1} f(i) r_2(x-i) + (a+b) \sum_{i=1}^{x} i f(i) F_{Z_N}(x-i), \tag{9.154}$$

with $r_1(0) = 0$.

There are also approaches to stabilize and safe guard this recursive algorithm against underflows and overflows, see Waldmann (1996, section 3) for details.

9.9.3 HIGHER ORDER RECURSIONS FOR DISCRETIZED SEVERITY DISTRIBUTIONS IN COMPOUND LDA MODELS

The Panjer recursion was introduced for the evaluation of the compound process distribution recursively when the frequency distribution had probabilities satisfying the recursive relationship given by

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1} \tag{9.155}$$

for some distribution $p \in \mathscr{P}_{10}$. In this section, recursions for compound distributions in which the counting distribution satisfies a generalized higher order recursion given by the most general representation

$$p_n = \sum_{i=1}^{k} \left(a_i + \sum_{j=0}^{l \vee i-1} \frac{b_{i,j}}{n-j}\right) p_{n-i}, \quad \forall n \in \{l+1, l+2, \dots\} \tag{9.156}$$

is considered, where the order is denoted by the 'lag' k. The results for the recursive evaluation of a compound distribution of an annual loss random variable $Z = \sum_{n=1}^{N} X_n$ with severity density $f \in \mathscr{P}_{10}$ and frequency distribution $p \in \mathscr{P}_{10}$ that also satisfies the higher order recursion in Equation 9.156 are given in Theorem 9.25 as derived in Sundt & Vernic (2009, chapter 5.1). Note that this results is the most general formulation that also incorporates cases in which $j = 0$ for which the recursion for the probabilities in the frequency distribution satisfy the simplified higher order recursion given by

$$p_n = \sum_{i=1}^{k} \left(a_i + \frac{b_i}{n} \right) p_{n-i}. \tag{9.157}$$

Theorem 9.25 *Consider the compound process LDA model with discretized severity distribution $f \in \mathscr{P}_{10}$ and frequency distribution $p \in \mathscr{P}_{10}$ that satisfies the recursive evaluation of the probability of $\Pr[N = n] = p_n$ given as a linear combination of k terms corresponding to the probabilities $\{\Pr[N = n-1], \ldots, \Pr[N = n-k]\}$ according to*

$$p_n = \sum_{i=1}^{k} \left(a_i + \sum_{j=0}^{l \vee i-1} \frac{b_{i,j}}{n-j} \right) p_{n-i}, \quad \forall n \in \{l+1, l+2, \ldots\}. \tag{9.158}$$

Then the compound process LDA annual loss distribution $f_Z(x)$ is recursively evaluated according to

$$\begin{aligned} f_Z(x) = {} & \frac{1}{1-\tau_a(f(0))} \sum_{n=1}^{l} \left(p_n - \sum_{i=1}^{k} a_i p_{n-i} \right) f^{(n)*}(x) \\ & - \frac{1}{1-\tau_a(f(0))} \sum_{i=1}^{k} \sum_{j=0}^{l \vee i-1} \sum_{n=j+1}^{l} \frac{b_{i,j}}{n-j} p_{n-i} f^{(n)*}(x) \\ & + \frac{1}{1-\tau_a(f(0))} \sum_{i=1}^{k} \sum_{y=1}^{x} a_i f^{(i)*}(y) f_Z(x-y) \\ & + \frac{1}{1-\tau_a(f(0))} \sum_{i=1}^{k} \sum_{j=0}^{l \vee i-1} b_{i,j} \left(c_{i-j} f^{(j)*}(x) \right. \\ & \left. + \frac{1}{i-j} \sum_{y=0}^{x-1} f^{(j)*}(y) \sum_{z=1}^{x-y} \frac{z}{x-y} f^{(i-j)*}(z) f_Z(x-y-z) \right), \end{aligned} \tag{9.159}$$

for all $x \in \{1, 2, \ldots\}$, $\tau_a(f(0))$ denoting the probability-generating function and coefficients

$$c_i = \sum_{n=1}^{\infty} \frac{1}{n} p_{n-i} f(0)^n, \quad \forall i \in \{1, 2, \ldots\}. \tag{9.160}$$

Clearly, the evaluation of the coefficients c_i proves an intractable quantity computationally due to the infinite summation. However, as noted in Sundt & Vernic (2009), if one further assumes that the severity distribution f is in the class $\mathscr{P}_{11}$, then the simplified result in Corollary 9.3 applies.

Corollary 9.3 *Consider the compound process LDA model with discretized severity distribution $f \in \mathscr{P}_{11}$ and frequency distribution $p \in \mathscr{P}_{10}$ that satisfies the recursive evaluation of the probability of $\Pr[N = n] = p_n$ given as a linear combination of k terms corresponding to the probabilities*

$\{\Pr[N = n-1], \ldots, \Pr[N = n-k]\}$ *according to*

$$p_n = \sum_{i=1}^{k} \left(a_i + \sum_{j=0}^{l \vee i-1} \frac{b_{i,j}}{n-j} \right) p_{n-i}, \quad \forall n \in \{l+1, l+2, \ldots\}. \tag{9.161}$$

Then the compound process LDA annual loss distribution $f_Z(x)$ *is recursively evaluated according to*

$$\begin{aligned} f_Z(x) = &\sum_{n=1}^{l} \left(p_n - \sum_{i=1}^{k} a_i p_{n-i} \right) f^{(n)*}(x) - \sum_{i=1}^{k} \sum_{j=0}^{l \vee i-1} \sum_{n=j+1}^{l} \frac{b_{i,j}}{n-j} p_{n-i} f^{(n)*}(x) \\ &+ \sum_{i=1}^{k} \sum_{j=0}^{l \vee i-1} \frac{b_{i,j}}{i-j} \sum_{y=j}^{x-1} f^{(j)*}(y) \sum_{z=i-j}^{x-y} \frac{z}{x-y} f^{(i-j)*}(z) f_Z(x-y-z) \\ &+ \sum_{i=1}^{k} \sum_{y=1}^{x} a_i f^{(i)*}(y) f_Z(x-y), \end{aligned} \tag{9.162}$$

for all $x \in \{1, 2, \ldots\}$.

9.9.4 RECURSIONS FOR DISCRETIZED SEVERITY DISTRIBUTIONS IN COMPOUND MIXED POISSON LDA MODELS

In this section, we consider generalizing the class of allowable frequency distributions to be mixed types. In particular, we briefly detail recursions for the evaluation of compound mixed Poisson distributions, the Wilmot class of mixing distributions and a simplified recursion. Then a recent generalization of the Panjer recursion because of Gerhold *et al.* (2010) will be considered which can be shown to improve numerical stability of the Panjer recursion and extend the application of the recursion to the class mixed-type compound distributions.

In this section, we consider the class of compound process distributions in which the frequency distribution is of a mixed type, generically represented according to Definition 9.11. Note that these distributions are also known as doubly stochastic processes and the Cox processes and have been utilized in numerous applications in the risk and insurance literature, see examples in OpRisk in Peters *et al.* (2011) and in the Bayesian context in Peters *et al.* (2009a).

Definition 9.11 (Mixed Poisson Type Frequency Distributions) *Consider* Λ *as a positive random variable with distribution* U. *Then define the frequency distribution for the number of losses* N *annually in the single risk LDA model to be defined conditionally as follows:*

$$\begin{aligned} &\Pr[N = n | \Lambda = \lambda] = \frac{\lambda^n}{n!} \exp(-\lambda), \quad \text{and,} \\ &\Pr[N = n] = \int_0^\infty \frac{\lambda^n}{n!} \exp(-\lambda) dU(\lambda) \quad \forall n \in \{0, 1, 2, \ldots\}. \end{aligned} \tag{9.163}$$

We note that several examples have been provided for such frequency distributions in the chapters on closed-form LDA models and insurance models in Chapters 6 and Cruz *et al.* (2015, chapter 16). It is also worth noting that the convolution between a finite number of mixed Poisson distributions is equivalent to considering a mixed Poisson distribution in which

the mixing distribution is obtained by the convolution between the mixing distributions of these distributions.

Next we detail how to perform recursive evaluation of the mixed Poisson compound distribution with severity distribution f and mixing distribution generically denoted by U given by

$$\begin{aligned} f_{Z_N}(x) &= \sum_{n=0}^{\infty} \mathbb{P}\mathrm{r}[N=n] f^{(n)*}(x) \\ &= \sum_{n=0}^{\infty} \int_0^{\infty} \frac{\lambda^n}{n!} \exp(-\lambda) dU(\lambda) f^{(n)*}(x). \end{aligned} \tag{9.164}$$

If one considers discretizing the severity distribution $f(x)$ to take support, without loss of generality, on the non-negative integers, then a recursive evaluation of the mixed Poisson compound process, due to Sundt & Jewell (1981, section 3.3) proceeds to evaluate $f_{Z_N}(x)$ for all $x \in \{0, 1, 2 \ldots\}$ as detailed in Proposition 9.6.

To understand this recursion, it is beneficial to first consider the basic recursion that one obtains when considering specifically the evaluation of a Poisson compound distribution in a single risk LDA model. As detailed in Sundt & Jewell (1981, theorem 2.2), one can efficiently evaluate the compound Poisson distribution according to Theorem 9.22 presented previously. Given this recursion in the context of the mixed Poisson compound distributions, this recursion can be turned into an alternative recursion with respect to moments of the Laplace transform of the mixing distribution. Consider the recursion

$$f_{Z_N}(x) = \frac{\lambda}{x} \sum_{y=1}^{x} y f(y) f_{Z_N}(x-y), \quad \forall x \in \{1, 2, \ldots\}, \tag{9.165}$$

and then multiply this by a power of the mixing random variable and its distribution, $\lambda^i dU(\lambda)$ to obtain in the notation of Sundt & Jewell (1981) the recursion for the ith integer moment

$$\begin{aligned} \nu_i(x) = \int_0^{\infty} \lambda^i f_{Z_N}(x;\lambda) dU(\lambda) = \frac{1}{x} \sum_{y=1}^{x} y f(y) \nu_{i+1}(x-y), \\ \forall x \in \{1, 2, \ldots\} \text{ and } \forall i \in \{0, 1, 2, \ldots\}. \end{aligned} \tag{9.166}$$

This second recursion for the values of $\nu_i(x)$ can then be utilized to evaluate mixed Poisson Compound distribution as detailed in Proposition 9.6.

Proposition 9.6 (Recursive Evaluation of Mixed Poisson Compound Distributions) *Consider a mixed Poisson compound distribution for a single risk LDA model, with discretized severity density given by $f(x)$ for $x \in \{0, 1, 2, \ldots\}$ and mixed frequency distribution in Definition 9.11. The evaluation of the compound process annual loss distribution at a point x according to*

$$f_{Z_N}(x) = \sum_{n=0}^{\infty} \int_0^{\infty} \frac{\lambda^n}{n!} \exp(-\lambda) dU(\lambda) f^{(n)*}(x), \tag{9.167}$$

proceeds by evaluation of the initialization $f_{Z_N}(0)$ *according to the Laplace transform of the mixing distribution* U *evaluated at* $1 - f(0)$ *via*

$$\nu_0(0) = f_{Z_N}(0) = \mathscr{L}[U(1 - f(0))] = \int_0^\infty \exp(-\lambda(1 - f(0)))\, dU(\lambda). \tag{9.168}$$

Then for all values of $y \in \{1, 2, \ldots, x\}$*, evaluate* y*-th derivative of the Laplace transform of the mixing distribution at* $1 - f(0)$ *according to*

$$\begin{aligned}\nu_y(0) = (-1)^y = \mathscr{L}[\lambda^y U(1 - f(0))] &= \int_0^\infty \lambda^y \exp(-\lambda(1 - f(0)))\, dU(\lambda) \\ &= (-1)^y \frac{d^y}{ds^y} \mathscr{L}[U(s)]|_{s=(1-f(0))}. \end{aligned} \tag{9.169}$$

Then for all $y \in \{1, 2, \ldots, x\}$ *and all* $z \in \{1, 2, \ldots, y\}$*, evaluate* $\nu_{y-z}(z)$ *via the recursion*

$$\begin{aligned}\nu_i(x) = \int_0^\infty \lambda^i f_{Z_N}(x; \lambda)\, dU(\lambda) = \frac{1}{x} \sum_{y=1}^{x} y f(y) \nu_{i+1}(x - y), \\ \forall x \in \{1, 2, \ldots\} \text{ and } \forall i \in \{0, 1, 2, \ldots\}. \end{aligned} \tag{9.170}$$

Finally, the value of the mixed Poisson compound distribution of $f_{Z_N}(y) = \nu_0(y)$.

Remark 9.16 *The computational efficiency and storage requirements for the recursive evaluation presented in Proposition 9.6 will become untenable as the value of* $x \to \infty$*. Unfortunately, this is precisely the situation in which one wishes to consider the utilization of this recursion in many applications related to estimation of the tail of the compound distribution for a single risk LDA model.*

Corollary 9.4 *Consider the severity distribution* f *which is discretized and takes values with positive probability only on the non-negative integers. Furthermore, assume that there exist functions* η *and* χ *on* $\{1, 2, \ldots, r\}$ *and an integer* r *either a positive integer or infinity that satisfies the first derivative of the power series representation of the probability mass function*

$$\frac{d}{ds}\mathbb{E}[s^X] = \int_0^\infty s^x\, dF(x) = \frac{\sum_{y=1}^{r} \eta(y) s^{y-1}}{1 - \sum_{y=1}^{r} \chi(y) s^y}, \tag{9.171}$$

then in the recursive evaluation of the mixed compound process, the stage that considers for all $y \in \{1, 2, \ldots, x\}$ *and all* $z \in \{1, 2, \ldots, y\}$ *evaluate* $\nu_{y-z}(z)$ *the recursive evaluation of*

$$\begin{aligned}\nu_i(x) = \int_0^\infty \lambda^i f_{Z_N}(x; \lambda)\, dU(\lambda) = \frac{1}{x} \sum_{y=1}^{x} y f(y) \nu_{i+1}(x - y), \\ \forall x \in \{1, 2, \ldots\} \text{ and } \forall i \in \{0, 1, 2, \ldots\}, \end{aligned} \tag{9.172}$$

can be replaced with an alternative recursion involving

$$\begin{aligned}\nu_i(x) = \sum_{y=1}^{r} \left(\frac{\eta(y)}{x} \nu_{i+1}(x - y) + \left(1 - \frac{y}{x}\right) \chi(y) \nu_i(x - y) \right), \\ \forall x \in \{1, 2, \ldots\} \text{ and } \forall i \in \{0, 1, 2, \ldots\}. \end{aligned} \tag{9.173}$$

It is well known that one can improve the efficiency of the recursion presented in Proposition 9.6 if additional restrictions on the mixing distribution are satisfied. The class of Wilmot mixing distributions correspond to an important set of such mixing distributions that allow for improved computational efficiency of the evaluation of mixed compound Poisson distributions, as defined in Definition 9.12, see numerous properties and details of this class in (Sundt & Vernic 2009, section 3.7).

Definition 9.12 (Wilmot Class of Mixing Distributions) *Consider a continuous mixing distribution U defined according to Definition 9.11. Furthermore, assume it takes a finite support on the interval $[\lambda_{\min}, \lambda_{\max}]$ with $\lambda_{\min} \geq 0$ and $\lambda_{\max} \leq \infty$. A mixing distribution $U(\lambda)$ on this support that admits a density $u(\lambda)$ that satisfies the condition that on the log scale, its derivative can be represented according to*

$$\frac{d}{d\lambda} \ln u(\lambda) = \frac{\sum_{i=0}^{k} \eta(i)\lambda^i}{\sum_{i=0}^{k} \chi(i)\lambda^i}, \forall \lambda \in [\lambda_{\min}, \lambda_{\max}], \tag{9.174}$$

belongs to the Wilmot Class of mixing distributions for some functions η and χ.

If one has a mixed-type frequency distribution model satisfying membership of the Wilmot class, then efficient recursions for the evaluation of the compound process with discretized severity proceed as detailed in Theorem 9.26, see derivations in Sundt & Vernic (2009, theorem 3.2).

Theorem 9.26 (Mixed Frequency Compound Recursion: Wilmot Class) *Consider a Poisson mixed frequency distribution in which the random intensity Λ is a positive random variable with distribution U. Then define the frequency distribution for the number of losses N annually in the single risk LDA model to be defined conditionally as follows:*

$$\begin{aligned} \mathbb{P}\mathrm{r}[N = n|\Lambda = \lambda] &= \frac{\lambda^n}{n!} \exp(-\lambda), \quad \text{and,} \\ \mathbb{P}\mathrm{r}[N = n] &= \int_0^\infty \frac{\lambda^n}{n!} \exp(-\lambda) dU(\lambda) \quad \forall n \in \{0, 1, 2, \ldots\}. \end{aligned} \tag{9.175}$$

Furthermore, assume the mixing distribution U is in the Wilmot class and satisfies that it takes support on the interval $[\lambda_{\min}, \lambda_{\max}]$ with $0 \leq \lambda_{\min} < \lambda_{\max} \leq \infty$ and admits a differentiable density u satisfying the condition for the Wilmot class membership

$$\frac{d}{d\lambda} \ln u(\lambda) = \frac{\sum_{i=0}^{k} \eta(i)\lambda^i}{\sum_{i=0}^{k} \chi(i)\lambda^i}, \forall \lambda \in [\lambda_{\min}, \lambda_{\max}]. \tag{9.176}$$

Then one has the following recursion for the annual loss distribution f_Z given the discretized severity distribution f_X over the non-negative integers $x \in \{1, 2, \ldots\}$ according to

$$\rho(k)\nu_k(x) = \sum_{y=1}^{x} f_X(y) \sum_{i=0}^{k} \chi(i)\nu_i(x-y) - \sum_{i=0}^{k-1} \rho(i)\nu_i(x) + w_{\lambda_{\min}}(x) - w_{\lambda_{\max}}(x), \tag{9.177}$$

with $\chi(-1) = \nu(-1) = \chi(k+1) = 0$ *and*

$$\rho(i) = (1 - f_X(0))\chi(i) - \nu(i) - (i+1)\chi(i+1), \quad i = -1, 0, 1, \ldots, k,$$

$$w_\lambda(x) = f_Z(x;\lambda)u(\lambda)\sum_{i=0}^{k}\chi(i)\lambda^i, \quad x = 0, 1, 2, \ldots; \lambda \in (\lambda_{\min}, \lambda_{\max}),$$

$$\nu_i(x) := \int \lambda^i f_Z(x;\lambda)dU(\lambda) = \frac{1}{x}\sum_{y=1}^{x} y f_X(y)\nu_{i+1}(x-y).$$

Next we consider the cases of recursions for compound distributions in which the severity distribution is not assumed to be discretized.

9.10 Continuous Versions of the Panjer Recursion

Although, in some settings, the discretization of a continuous severity distribution might be justifiable; this is not the preferred approach in most OpRisk models. However, the Panjer recursion approach may also be applied in a continuous setting, leading to the recursion for the density given by

$$f_{Z_N}(x) = p_1 f_X(x) + \int_0^x \left(a + \frac{b\tau}{x}\right) f_X(\tau) f_{Z_N}(x-\tau)d\tau, \tag{9.178}$$

where $f_X(x)$ is the severity density and a, b and $p_1 = \mathbb{P}\mathrm{r}[N = 1]$ parameterize the frequency distribution of the compound process. There are many approaches to evaluate this expression. We focus primarily on importance sampling-based approaches in this chapter. For alternative numerical techniques to the Panjer recursion, see works such as inversion transforms (fast Fourier transforms) and series expansions (Bergstrom 1953); see Shevchenko (2011) and Cruz (2002) and references therein.

9.10.1 THE PANJER RECURSION VIA VOLTERRA INTEGRAL EQUATIONS OF THE SECOND KIND

One can now observe that the Panjer recursion specified in Equation 9.178 can be recognized as the Volterra equation of the second kind, see discussions in Peters *et al.* (2007) and references therein. In general, the Volterra integral equation of the second kind takes the form:

$$f(x) = g(x) + \int_0^x K(x, x_1, f(x_1))dx_1. \tag{9.179}$$

Therefore, one can observe that for the Panjer recursion if one selects a linear Volterra equation in which,

$$K(x, x_1, f(x_1)) = k(x, x_1)f(x_1), \tag{9.180}$$

this will produce

$$f(x) = g(x) + \int_0^x k(x, x_1)f(x_1)dx_1, \tag{9.181}$$

allowing one to make an association directly between the linear Volterra equation of the second kind and the Panjer recursion. This involves making the following identifications:

$$\begin{aligned} x_1 &= x - \tau, \\ g(x) &= p_1 f_X(x), \\ k(x, x_1) &= \left(a + b\frac{x - x_1}{x} \right) f_X(x - x_1), \text{ and} \\ f(x_1) &= f_{Z_N}(x_1). \end{aligned}$$

Therefore, one can obtain the following recursive representation for this integral equation, first working with the representation from the Volterra integral equation,

$$\begin{aligned} f(x) &= g(x) + \int_0^x k(x, x_1) f(x_1) dx_1 \\ &= g(x) + \int_0^x k(x, x_1) \left[g(x) + \int_0^{x_1} k(x_1, x_2) f(x_2) dx_2 \right] dx_1, \end{aligned} \tag{9.182}$$

where $g : [0, x] \mapsto \mathbb{R}$ and $k : [0, x] \times [0, x] \mapsto \mathbb{R}$ are known functions and $f : [0, x] \mapsto \mathbb{R}$ is unknown. Now if one recognizes that this equation is also expressed as according to the expression

$$f(x) = g(x) + \int_0^x r(x, x_1) g(x_1) dx_1, \tag{9.183}$$

where r is the resolvent kernel for the Volterra equation of the second kind which can, therefore, be expressed according to a von Neumann series expansion, see Baker (2000). This produces

$$r(x, x_1) = \sum_{n=1}^{\infty} k^n(x, x_1), \tag{9.184}$$

where one defines $k^0(x, x_1) = 1$ $k^1(x, x_1) = k(x, x_1)$ and

$$k^n(x, x_1) = \int_0^x k(x, u) k^{n-1}(u, x_1) du, \quad n = 2, 3, 4, \ldots, \tag{9.185}$$

such that the von Neumann series expansion exists under the condition that

$$\sum_{n=0}^{\infty} \int_0^x |k^n(x_0, x_n| dx_n < \infty. \tag{9.186}$$

Therefore, if one applies this series expansion to Equation 9.181 , one obtains for any point $x_0 \in \mathbb{R}^+$,

$$f(x_0) = g(x_0) + \sum_{n=1}^{\infty} \int_0^{x_0} \cdots \int_0^{x_{n-1}} g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l) dx_{1:n}, \tag{9.187}$$

with notation $x_{1:n}$ denoting an n-tuple $(x_1, x_2, \ldots, x_n)$. Now as in Peters *et al.* (2007), we define the domains of integration according to the following notations where $D_k(x_{k-1}) =$

$[0, x_{k-1}]$ is the conditional one-dimensional domains of integration and the domain of integration of the nth term in the summation as

$$D_{1:n}(x_0) = \{(x_1, \ldots, x_n) : x_0 > x_1 > \cdots > x_n\},$$

with the convention that $D_{1:0}(x_0) = \{\emptyset\}$. Under this representation, one may rewrite the series expansion according to

$$f(x_0) = g(x_0) + \sum_{n=1}^{\infty} \int_{D_{1:n}(x_0)} g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l) dx_{1:n}. \tag{9.188}$$

Additionally, we define the domain $\widehat{D}_{0:n}(D_0) = \{(x_0, x_1, \ldots, x_n) : D_0 \ni x_0 > x_1 > \cdots > x_n\}$ and D_0 corresponds to the region of values over which one wishes to characterize the annual loss distribution, for example, an interval $[x_a, x_b]$.

Clearly, one can now start to characterize the tail distribution of the compound process under this framework according to the integral over the domain $[x, \infty)$ using the series representation given by

$$\begin{aligned} \overline{F}_{Z_N}(x) &= \mathbb{P}\mathrm{r}[Z_N > x] = \sum_{n=1}^{\infty} \mathbb{P}\mathrm{r}[N = n] \overline{F}_{Z_n}(x) \\ &= \int_x^{\infty} g(\tau) d\tau + \sum_{n=1}^{\infty} \int_x^{\infty} \int_{D_{1:n}(\tau)} g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l) dx_{1:n} d\tau. \end{aligned} \tag{9.189}$$

In Peters *et al.* (2007), a set of importance sampling estimation methodologies were developed to approximation numerically the recursive integral representations developed. This involved primarily importance sampling-based methods on a path space. The choice of the importance sampling distribution was selected in two settings, one for efficiency and one utilizing a trans-dimensional MCMC proposal which was provably optimal in minimizing the variance of the importance sampling weights.

9.10.2 IMPORTANCE SAMPLING SOLUTIONS TO THE CONTINUOUS PANJER RECURSION

To introduce the importance sampling estimation framework to estimate the compound process distribution pointwise or over an interval, it will be convenient to introduce the following additional notation:

$$\begin{aligned} f_0(x_0) &= g(x_0), \quad \text{and} \\ f_n(x_{0:n}) &= g(x_n) \prod_{l=1}^{n} k(x_{l-1}, x_l), \end{aligned} \tag{9.190}$$

which then allows for a representation of the evaluation of the density either at a point x_0 by

$$f(x_0) = f_0(x_0) + \sum_{n=1}^{\infty} \int_{D_{1:n}(x_0)} f_n(x_{0:n}) dx_{1:n}, \tag{9.191}$$

or on an interval D_0 using the $\widehat{D}_{0:n}(D_0)$.

Therefore, one can frame this quantity of interest according to an expectation with respect to some importance sampling distribution π according to

$$\begin{aligned} f(x) &= \frac{f_0(x)}{\pi(0)} + \sum_{n=1}^{\infty} \int_{D_{1:n}(x)} \frac{f_n(x, x_{1:n})}{\pi(n, x_{1:n})} \pi(n, x_{1:n}) dx_{1:n} \\ &= \mathbb{E}\left[\frac{f_n(x, x_{1:n})}{\pi(n, x_{1:n})} \right]. \end{aligned} \tag{9.192}$$

As discussed, there are then two estimation problems of interest, the estimation of $f(x)$ pointwise and the characterization of $f(x)$ over some interval by obtaining samples from its restriction to that interval.

The space on which the importance sampling is performed is now a path space because it corresponds to either

1. estimation of $f(x)$ pointwise via an importance sampling space $\cup_{n=0}^{\infty}\{n\} \times D_{1:n}(x)$, or
2. estimation of $f(x)$ over an interval D_0 via an importance sampling space $\cup_{n=0}^{\infty}\{n\} \times \widehat{D}_{1:n}([x_a, x_b])$.

As noted in Peters *et al.* (2007), when one is interested in estimating the function pointwise, as $f_0(x)$ is known, it would be more efficient in the sense that variance would be reduced on both a per sample basis and a per unit of computation basis to instead estimate $f(x) - f_0(x)$ by importance sampling on the smaller space $\cup_{n=1}^{\infty}\{n\} \times D_{1:n}(x)$, and this approach introduces no further complications.

It is critical to any importance sampling-based procedure to ensure a suitable importance sampling distribution is selected. In this section, we detail a simple Markov chain proposal formulation and refer the interested reader to a trans-dimensional Markov chain proposal developed to minimize the variance of the importance sampling weights in Peters *et al.* (2007, section 3.2).

Consider a proposal distribution that is intuitive to understand and simple to simulate from as detailed in Peters *et al.* (2007, section 3.1). The solution in this setting would involve starting with a Markov chain from x (or with some initial distribution μ which covers the region of interest if we wish to characterize f over some interval rather than at a point) and a transition kernel for the Markov chain denoted by $M(x, y)$ which is the probability density for going from state x to state y. The initial distribution μ, when it is used, and transition kernel M are selected such that $\mu(x) > 0$ over the region of interest and $M(x, y) > 0$ if $k(x, y) \neq 0$, which is important to ensure that the importance sampling scheme is well defined over the domain of interest, avoiding bias in estimates. In addition, the space explored by M is designed to have an absorbing cemetery state that we denote by d, where $d \notin [0, \infty)$ and $M(x, d) = P_d$ for any x. Therefore, the proposal we consider for the importance sampler over the path space in the case of considering a evaluation at a point x_0 takes the following form $\pi(n, x_{1:n}) = \pi(n)\pi_n(x_{1:n})$ with

$$\begin{aligned} \pi(n) &= \mathbb{P}\mathrm{r}[X_{1:n} \in D_{1:n}(x_0), X_{n+1} = \{d\}] = (1 - P_d)^n P_d, \\ \pi_n(x_{1:n}) &= \frac{\prod_{k=1}^{n} M(x_{k-1}, x_k)}{(1 - P_d)^n}. \end{aligned} \tag{9.193}$$

Note that the dependency of $\pi_n(x_{1:n})$ on the point x_0 is not made explicit here but is understood from the construction.

Now using this proposal, one can develop an importance sampling-based estimation. To present this estimation, we introduce the particle notation in which we represent the annual loss distribution $f_{Z_N}(x)$ according to an empirical measure with P independent path samples, either at a point $x = x_0$ according to Equation 9.194

$$\widehat{f}_{Z_N}(x_0) = \frac{1}{P}\sum_{i=1}^{P} W(x_0, X^{(i)}_{1:n^{(i)}}), \tag{9.194}$$

or over an interval $D_0 = [x_a, x_b]$ according to Equation 9.195

$$\widehat{f}_{Z_N}(x_0) = \frac{1}{P}\sum_{i=1}^{P} W(X^{(i)}_{0:n^{(i)}})\delta(x_0 - X^{(i)}_0), \tag{9.195}$$

where for the i-th particle (independent sample), we denote the path with $n^{(i)}$ stages, run until absorption, by $X^{(i)}_{0:n^{(i)}}$, the importance weight for the path by $W(X^{(i)}_{0:n^{(i)}})$ and the Dirac mass $\delta(x_0 - X^{(i)}_0)$ located at $X^{(i)}_0$.

The importance sampling approximation of the annual loss density $f_{Z_N}(x)$ is given by the following steps:

- Generate P independent Markov chain paths $\{X^{(i)}_{0:n^{(i)}+1}\}_{i=1}^{P}$ until absorption $X^{(i)}_{n^{(i)}+1} = d$.
- Evaluate the importance weights for each particle on the path space. If evaluation of the annual loss density at a point is desired for a value x_0, then the weight is given by

$$W(X^{(i)}_{0:n^{(i)}}) = \begin{cases} \left(\prod_{s=1}^{n^{(i)}} \dfrac{k\left(X^{(i)}_{s-1}, X^{(i)}_s\right)}{M(X^{(i)}_{s-1}, X^{(i)}_s)}\right) \dfrac{g(X^{(i)}_{n^{(i)}})}{P_d}, & \text{if } n^{(i)} \geq 1, \\ \dfrac{g(X^{(i)}_0)}{\mu(X^{(i)}_0)P_d}, & \text{if } n^{(i)} = 0. \end{cases} \tag{9.196}$$

Whilst, if X_0 is being sampled from some distribution μ in order to characterize f over some interval, then the importance weight function becomes

$$W(X^{(i)}_{0:n^{(i)}}) = \begin{cases} \dfrac{1}{\mu\left(X^{(i)}_0\right)} \left(\prod_{n=1}^{n^{(i)}} \dfrac{k\left(X^{(i)}_{n-1}, X^{(i)}_n\right)}{M(X^{(i)}_{n-1}, X^{(i)}_n)}\right) \dfrac{g(X^{(i)}_{n^{(i)}})}{P_d}, & \text{if } n^{(i)} \geq 1, \\ \dfrac{g(X^{(i)}_0)}{\mu(X^{(i)}_0)P_d}, & \text{if } n^{(i)} = 0. \end{cases} \tag{9.197}$$

Then the empirical measure

$$\widehat{f}_Z(x_0) = \frac{1}{N}\sum_{i=1}^{N} W_1(X^{(i)}_{0:n^{(i)}})\delta(x_0 - X^{(i)}_0),$$

forms an unbiased Monte Carlo approximation of the expectation of $f_Z(z)$ for any set D_0 given by $\mathbb{E}\left[\int_{D_0}\widehat{f}(x_0)dx_0\right] = \int_{D_0} f(x_0)dx_0$. Furthermore, detailed discussions on the optimal choice with respect to minimizing the variance of the importance weights is developed in Peters *et al.* (2007) and Doucet *et al.* (2010).

APPENDIX A

Miscellaneous Definitions and List of Distributions

A.1 Indicator Function

The often used indicator symbol $\mathbb{I}_{\{\cdot\}}$ is defined as

$$\mathbb{I}_{\{\cdot\}} = \begin{cases} 1, & \text{if condition in } \{\cdot\} \text{ is true,} \\ 0, & \text{otherwise.} \end{cases} \tag{A.1}$$

In addition on occasion we will also utilise $\mathbb{I}[\cdot]$.

A.2 Gamma Function

The standard gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt, \ \alpha > 0. \tag{A.2}$$

A.3 Discrete Distributions

A.3.1 POISSON DISTRIBUTION

A Poisson distribution function is denoted as $Poisson(\lambda)$. The random variable N has a Poisson distribution, denoted as $N \sim Poisson(\lambda)$, if its probability mass function is

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

$$p(k) = \mathbb{P}\mathrm{r}[N = k] = \frac{\lambda^k}{k!} e^{-\lambda},\ \lambda > 0, \tag{A.3}$$

for all $k \in \{0, 1, 2, \dots\}$. Expectation, variance and variational coefficient of a random variable $N \sim Poisson(\lambda)$ are

$$\begin{aligned} \mathbb{E}[N] &= \lambda, \\ \mathbb{V}\mathrm{ar}[N] &= \lambda, \\ \mathrm{Vco}[N] &= \frac{1}{\sqrt{\lambda}}. \end{aligned}$$

A.3.2 BINOMIAL DISTRIBUTION

The Binomial distribution function is denoted as $Binomial(n, p)$. The random variable N has a Binomial distribution, denoted as $N \sim Binomial(n, p)$, if its probability mass function is

$$p(k) = \mathbb{P}\mathrm{r}[N = k] = \binom{n}{k} p^k (1-p)^{n-k},\ p \in (0, 1),\ n \in 1, 2, \dots, \tag{A.4}$$

for all $k \in \{0, 1, 2, \dots, n\}$. Expectation, variance and variational coefficient of a random variable $N \sim Binomial(n, p)$ are

$$\begin{aligned} \mathbb{E}[N] &= np, \\ \mathbb{V}\mathrm{ar}[N] &= np(1-p), \\ \mathrm{Vco}[N] &= \sqrt{\frac{1-p}{np}}. \end{aligned}$$

Remark A.1 *N is the number of successes in n independent trials, where p is the probability of a success in each trial.*

A.3.3 NEGATIVE BINOMIAL DISTRIBUTION

A Negative Binomial distribution function is denoted as $NegBinomial(r, p)$. The random variable N has a Negative Binomial distribution, denoted $N \sim NegBinomial(r, p)$, if its probability mass function is

$$p(k) = \mathbb{P}\mathrm{r}[N = k] = \binom{r+k-1}{k} p^r (1-p)^k,\ p \in (0, 1),\ r \in (0, \infty), \tag{A.5}$$

for all $k \in \{0, 1, 2, \dots\}$. Here, the generalized binomial coefficient is

$$\binom{r+k-1}{k} = \frac{\Gamma(k+r)}{k!\Gamma(r)}, \tag{A.6}$$

where $\Gamma(r)$ is the gamma function.

Expectation, variance and variational coefficient of a random variable $N \sim NegBinomial(r,p)$ are

$$\mathbb{E}[N] = \frac{r(1-p)}{p},$$
$$\mathbb{V}\text{ar}[N] = \frac{r(1-p)}{p^2},$$
$$\text{Vco}[N] = \frac{1}{\sqrt{r(1-p)}}.$$

Remark A.2 *If r is an positive integer, N is the number of failures in a sequence of independent trials until r successes, where p is the probability of a success in each trial.*

A.3.4 DOUBLY STOCHASTIC POISSON PROCESS (COX PROCESS)

Let $(\Omega, F, \mathbb{P})$ be a probability space with information structure (filtration) given by $F = \{\mathscr{F}_t, t \in [0,T]\}$. Let N_t be a point process adapted to F Let λ_t be a non-negative process adapted to F such that

$$\int_0^t \lambda_s ds < \infty, \;\; a.s. \tag{A.7}$$

If for all $0 \le t_1 \le t_2$ and $u \in \mathbb{R}$ one has

$$\mathbb{E}[e^{\mathrm{i}u(N_{t_2} - N_{t_1})} | \mathscr{F}_{t_2}] = \exp\left\{ (e^{\mathrm{i}u} - 1) \int_{t_1}^{t_2} \lambda_s ds \right\} \tag{A.8}$$

then N_t is called a $\mathscr{F}_t$-doubly stochastic Poisson process with intensity λ_t where $\mathscr{F}_t = \sigma\{\lambda_s; s \le t\}$. One has the following probabilities

$$\mathbb{P}\text{r}[N_{t_2} - N_{t_1} = k | \lambda_s ds;\; t_1 \le s \le t_2] = \frac{\exp(-\int_{t_1}^{t_2} \lambda_s ds)[\int_{t_1}^{t_2} \lambda_s ds]^k}{k!} \tag{A.9}$$

and in addition one has for τ_k the length of time interval between the $(k-1)$-th and the k-th point the following distribution

$$\mathbb{P}\text{r}[\tau_k > t \,|\lambda_s;\; t_{k-1} \le s \le t_{k-1} + t] = \exp\left(- \int_{t_{k-1}}^{t_{k-1}+t} \lambda_s ds \right). \tag{A.10}$$

A.4 Continuous Distributions

A.4.1 UNIFORM DISTRIBUTION

A Uniform distribution function is denoted as $Uniform(a,b)$. The random variable X has a Uniform distribution, denoted as $X \sim Uniform(a,b)$, if its probability density function is

$$f(x) = \frac{1}{b-a}, \; a < b, \tag{A.11}$$

for $x \in [a, b]$. Expectation, variance and variational coefficient of a random variable $X \sim Uniform(a, b)$ are

$$\mathbb{E}[X] = \frac{a+b}{2},$$
$$\mathbb{V}\text{ar}[X] = \frac{(b-a)^2}{12},$$
$$\text{Vco}[X] = \frac{b-a}{\sqrt{3}(a+b)}.$$

A.4.2 NORMAL (GAUSSIAN) DISTRIBUTION

A Normal (Gaussian) distribution function is denoted as $Normal(\mu, \sigma^2)$. The random variable X has a Normal distribution, denoted as $X \sim Normal(\mu, \sigma^2)$, if its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \ \sigma^2 > 0, \ \mu \in \mathbb{R}, \tag{A.12}$$

for all $x \in \mathbb{R}$. Expectation, variance and variational coefficient of a random variable $X \sim Normal(\mu, \sigma^2)$ are

$$\mathbb{E}[X] = \mu,$$
$$\mathbb{V}\text{ar}[X] = \sigma^2,$$
$$\text{Vco}[X] = \sigma/\mu.$$

A.4.3 INVERSE GAUSSIAN DISTRIBUTION

An Inverse Gaussian distribution function is denoted as $InverseGaussian(\mu, \gamma)$. The random variable X has an Inverse Gaussian distribution, denoted $X \sim InverseGaussian(\mu, \gamma)$, if its probability density function is

$$f(x) = \left(\frac{\gamma}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\gamma(x-\mu)^2}{2\mu^2 x}\right), \ x > 0, \tag{A.13}$$

where parameters $\mu > 0$ and $\gamma > 0$. The corresponding distribution function is

$$F(x) = \Phi\left(\sqrt{\frac{\gamma}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \exp\left(\frac{2\gamma}{\mu}\right)\Phi\left(-\sqrt{\frac{\gamma}{x}}\left(\frac{x}{\mu} + 1\right)\right), \tag{A.14}$$

where $\Phi(\cdot)$ is the standard Normal distribution. Expectation and variance of $X \sim InverseGaussian(\mu, \lambda)$ are

$$\mathbb{E}[X] = \mathbb{V}\text{ar}[X] = \frac{\mu^3}{\gamma}.$$

If $X_1, \ldots, X_n$ are independent and $X_i \sim InverseGaussian(\mu w_i, \gamma {w_i}^2)$, then

$$S_n = \sum_{i=1}^{n} X_i \sim InverseGaussian\left(\mu \sum_{i=1}^{n} w_i, \gamma\left(\sum_{i=1}^{n} w_i\right)^2\right). \tag{A.15}$$

A.4.4 LogNORMAL DISTRIBUTION

A LogNormal distribution function is denoted as $LogNormal(\mu, \sigma^2)$. The random variable X has a LogNormal distribution, denoted as $X \sim LogNormal(\mu, \sigma^2)$, if its probability density function is

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right),\ \sigma^2 > 0,\ \mu \in \mathbb{R}, \tag{A.16}$$

for $x > 0$. Expectation, variance and variational coefficient of a random variable $X \sim LogNormal(\mu, \sigma^2)$ are

$$\begin{aligned}
\mathbb{E}[X] &= e^{\mu+\frac{1}{2}\sigma^2},\\
\mathbb{V}\mathrm{ar}[X] &= e^{2\mu+\sigma^2}(e^{\sigma^2}-1),\\
\mathrm{Vco}[X] &= \sqrt{e^{\sigma^2}-1}.
\end{aligned}$$

A.4.5 STUDENT'S *t*-DISTRIBUTION

A t-distribution function is denoted as $\mathcal{T}(\nu, \mu, \sigma^2)$. The random variable X has a t-distribution, denoted as $X \sim \mathcal{T}(\nu, \mu, \sigma^2)$, if its probability density function is

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \frac{1}{\sqrt{\nu\pi}} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2} \tag{A.17}$$

for $\sigma^2 > 0$, $\mu \in \mathbb{R}$, $\nu = 1, 2, \ldots$ and all $x \in \mathbb{R}$. Expectation, variance and variational coefficient of a random variable $X \sim \mathcal{T}(\nu, \mu, \sigma^2)$ are

$$\begin{aligned}
\mathbb{E}[X] &= \mu \text{ if } \nu > 1,\\
\mathbb{V}\mathrm{ar}[X] &= \sigma^2 \frac{\nu}{\nu-2} \text{ if } \nu > 2,\\
\mathrm{Vco}[X] &= \frac{\sigma}{\mu}\sqrt{\frac{\nu}{\nu-2}} \text{ if } \nu > 2.
\end{aligned}$$

A.4.6 GAMMA DISTRIBUTION

A Gamma distribution function is denoted as $Gamma(\alpha, \beta)$. The random variable X has a Gamma distribution, denoted as $X \sim Gamma(\alpha, \beta)$, if its probability density function is

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} \exp(-x/\beta),\ \alpha > 0,\ \beta > 0, \tag{A.18}$$

for $x > 0$. Expectation, variance and variational coefficient of a random variable $X \sim Gamma(\alpha, \beta)$ are

$$\begin{aligned}
\mathbb{E}[X] &= \alpha\beta,\\
\mathbb{V}\mathrm{ar}[X] &= \alpha\beta^2,\\
\mathrm{Vco}[X] &= 1/\sqrt{\alpha}.
\end{aligned}$$

A.4.7 WEIBULL DISTRIBUTION

A Weibull distribution function is denoted as $Weibull(\alpha, \beta)$. The random variable X has a Weibull distribution, denoted as $X \sim Weibull(\alpha, \beta)$, if its probability density function is

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp(-(x/\beta)^\alpha),\ \alpha > 0,\ \beta > 0, \tag{A.19}$$

for $x > 0$. The corresponding distribution function is

$$F(x) = 1 - \exp(-(x/\beta)^\alpha),\ \alpha > 0,\ \beta > 0. \tag{A.20}$$

Expectation and variance of a random variable $X \sim Weibull(\alpha, \beta)$ are

$$\begin{aligned} \mathbb{E}[X] &= \beta\Gamma(1 + 1/\alpha), \\ \mathbb{V}\text{ar}[X] &= \beta^2(\Gamma(1 + 2/\alpha) - (\Gamma(1 + 1/\alpha))^2). \end{aligned}$$

A.4.8 INVERSE CHI-SQUARED DISTRIBUTION

An Inverse Chi-squared distribution is denoted as $InvChiSq(\nu, \beta)$. The random variable X has an Inverse Chi-squared distribution, denoted as $X \sim InvChiSq(\nu, \beta)$, if its probability density function is

$$f(x) = \frac{(x/\beta)^{-1-\nu/2}}{\beta\Gamma(\nu/2)2^{\nu/2}} \exp\left(-\frac{\beta}{2x}\right) \tag{A.21}$$

for $x > 0$ and parameters $\nu > 0$ and $\beta > 0$. Expectation and variance of $X \sim InvChiSq(\nu, \beta)$ are

$$\begin{aligned} \mathbb{E}[X] &= \frac{\beta}{\nu - 2},\ \text{for}\ \nu > 2, \\ \mathbb{V}\text{ar}[X] &= \frac{2\beta^2}{(\nu - 2)^2(\nu - 4)}\ \text{for}\ \nu > 4. \end{aligned}$$

A.4.9 PARETO DISTRIBUTION (ONE PARAMETER)

A one-parameter Pareto distribution function is denoted as $Pareto(\xi, x_0)$. The random variable X has a Pareto distribution, denoted as $X \sim Pareto(\xi, x_0)$, if its distribution function is

$$F(x) = 1 - \left(\frac{x}{x_0}\right)^{-\xi}, \quad x \geq x_0, \tag{A.22}$$

where $x_0 > 0$ and $\xi > 0$. The support starts at x_0, which is typically known and not considered as a parameter. Therefore, the distribution is referred to as a single-parameter Pareto. The corresponding probability density function is

$$f(x) = \frac{\xi}{x_0}\left(\frac{x}{x_0}\right)^{-\xi-1}. \tag{A.23}$$

Expectation, variance and variational coefficient of $X \sim Pareto(\xi, x_0)$ are

$$\mathbb{E}[X] = x_0 \frac{\xi}{\xi - 1} \text{ if } \xi > 1,$$

$$\mathbb{V}\text{ar}[X^2] = x_0^2 \frac{\xi}{(\xi - 1)^2(\xi - 2)} \text{ if } \xi > 2,$$

$$\text{Vco}[X] = \frac{1}{\sqrt{\xi(\xi - 2)}} \text{ if } \xi > 2.$$

A.4.10 PARETO DISTRIBUTION (TWO PARAMETER)

A two-parameter Pareto distribution function is denoted as $Pareto_2(\alpha, \beta)$. The random variable X has a Pareto distribution, denoted as $X \sim Pareto_2(\alpha, \beta)$, if its distribution function is

$$F(x) = 1 - \left(1 + \frac{x}{\beta}\right)^{-\alpha}, \quad x \geq 0, \tag{A.24}$$

where $\alpha > 0$ and $\beta > 0$. The corresponding probability density function is

$$f(x) = \frac{\alpha \beta^\alpha}{(x + \beta)^{\alpha+1}}. \tag{A.25}$$

The moments of a random variable $X \sim Pareto_2(\alpha, \beta)$ are

$$\mathbb{E}[X^k] = \frac{\beta^k k!}{\prod_{i=1}^k (\alpha - i)}, \quad \alpha > k.$$

A.4.11 GENERALIZED PARETO DISTRIBUTION

A generalized Pareto distribution (GPD) distribution function is denoted as $GPD(\xi, \beta)$. The random variable X has a GPD distribution, denoted as $X \sim GPD(\xi, \beta)$, if its distribution function is

$$H_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x/\beta), & \xi = 0, \end{cases} \tag{A.26}$$

where $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$. The corresponding probability density function is

$$h(x) = \begin{cases} \frac{1}{\beta}(1 + \xi x/\beta)^{-\frac{1}{\xi}-1}, & \xi \neq 0, \\ \frac{1}{\beta} \exp(-x/\beta), & \xi = 0. \end{cases} \tag{A.27}$$

Expectation, variance and variational coefficient of $X \sim GPD(\xi, \beta)$, $\xi \geq 0$, are

$$\mathbb{E}[X^n] = \frac{\beta^n n!}{\prod_{k=1}^n (1 - k\xi)}, \ \xi < \frac{1}{n};$$

$$\mathbb{E}[X] = \frac{\beta}{1 - \xi}, \ \xi < 1;$$

$$\mathbb{V}\text{ar}[X^2] = \frac{\beta^2}{(1 - \xi)^2(1 - 2\xi)},$$

$$\mathrm{Vco}[X] = \frac{1}{\sqrt{1-2\xi}},\ \xi < \frac{1}{2}.$$

A.4.12 BETA DISTRIBUTION

A Beta distribution function is denoted as $Beta(\alpha, \beta)$. The random variable X has a beta distribution, denoted as $X \sim Beta(\alpha, \beta)$, if its probability density function is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \le x \le 1, \tag{A.28}$$

for $\alpha > 0$ and $\beta > 0$. Expectation, variance and variational coefficient of a random variable $X \sim Beta(\alpha, \beta)$ are

$$\mathbb{E}[X] = \frac{\alpha}{\alpha+\beta},$$
$$\mathbb{V}\mathrm{ar}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)},$$
$$\mathrm{Vco}[X] = \sqrt{\frac{\beta}{\alpha(1+\alpha+\beta)}}.$$

A.4.13 GENERALIZED INVERSE GAUSSIAN DISTRIBUTION

A generalized Inverse Gaussian (GIG) distribution function is denoted as $GIG(\omega, \phi, \nu)$. The random variable X has a GIG distribution, denoted as $X \sim GIG(\omega, \phi, \nu)$, if its probability density function is

$$f(x) = \frac{(\omega/\phi)^{(\nu+1)/2}}{2K_{\nu+1}(2\sqrt{\omega\phi})} x^{\nu} e^{-x\omega - x^{-1}\phi}, \quad x > 0, \tag{A.29}$$

where $\phi > 0$, $\omega \ge 0$ if $\nu < -1$; $\phi > 0$, $\omega > 0$ if $\nu = -1$; $\phi \ge 0$, $\omega > 0$ if $\nu > -1$; and

$$K_{\nu+1}(z) = \frac{1}{2}\int_0^{\infty} u^{\nu} e^{-z(u+1/u)/2} \mathrm{d}u.$$

$K_{\nu}(z)$ is called a modified Bessel function of the third kind; see, for instance, Abramowitz & Stegun (1965, p. 375).

The moments of a random variable $X \sim GIG(\omega, \phi, \nu)$ are not available in a closed form through elementary functions but can be expressed in terms of Bessel functions:

$$\mathbb{E}[X^{\alpha}] = \left(\frac{\phi}{\omega}\right)^{\alpha/2} \frac{K_{\nu+1+\alpha}(2\sqrt{\omega\phi})}{K_{\nu+1}(2\sqrt{\omega\phi})}, \quad \alpha \ge 1,\ \phi > 0,\ \omega > 0.$$

Often, using notation $R_{\nu}(z) = K_{\nu+1}(z)/K_{\nu}(z)$, it is written as

$$\mathbb{E}[X^{\alpha}] = \left(\frac{\phi}{\omega}\right)^{\alpha/2} \prod_{k=1}^{\alpha} R_{\nu+k}(2\sqrt{\omega\phi}), \quad \alpha = 1, 2, \ldots$$

The mode is easily calculated from $\frac{\partial}{\partial x} x^{\nu} e^{-(\omega x + \phi/x)} = 0$ as

$$\text{mode}(X) = \frac{1}{2\omega}(\nu + \sqrt{\nu^2 + 4\omega\phi}),$$

that differs only slightly from the expected value for large ν, that is,

$$\text{mode}(X) \to \mathbb{E}[X] \qquad \text{for,} \quad \nu \to \infty.$$

A.4.14 d-VARIATE NORMAL DISTRIBUTION

A d-variate normal distribution function is denoted as $Normal(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1, \ldots \mu_d)^T \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}$ is a positive definite matrix $(d \times d)$. The corresponding probability density function is

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{det\, \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right), \ \boldsymbol{x} \in \mathbb{R}^d, \tag{A.30}$$

where $\boldsymbol{\Sigma}^{-1}$ is the inverse of the matrix $\boldsymbol{\Sigma}$. Expectations and covariances of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^T \sim Normal(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are

$$\mathbb{E}[X_i] = \mu_i, \ \mathbb{C}\text{ov}[X_i, X_j] = \Sigma_{i,j}, \ i, j = 1, \ldots, d. \tag{A.31}$$

A.4.15 d-VARIATE t-DISTRIBUTION

A d-variate t-distribution function with ν degrees of freedom is denoted as $\mathcal{T}_d(\nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\nu > 0$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_d)^T \in \mathbb{R}^d$ is a location vector and $\boldsymbol{\Sigma}$ is a positive definite matrix $(d \times d)$. The corresponding probability density function is

$$f(\boldsymbol{x}) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{(\nu\pi)^{d/2}\Gamma\left(\frac{\nu}{2}\right)\sqrt{det\, \boldsymbol{\Sigma}}} \left(1 + \frac{(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+d}{2}}, \tag{A.32}$$

where $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}^{-1}$ is the inverse of the matrix $\boldsymbol{\Sigma}$. Expectations and covariances of a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^T \sim Normal(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are

$$\mathbb{E}[X_i] = \mu_i, \text{ if } \nu > 1, \ i = 1, \ldots, d;$$
$$\mathbb{C}\text{ov}[X_i, X_j] = \nu\Sigma_{i,j}/(\nu - 2), \text{ if } \nu > 2, \ i, j = 1, \ldots, d.$$

References

Aaronson, J., & Denker, M. 1998. Characteristic functions of random variables attracted to 1-stable laws. *Annals of Probability*, **26**(1), 399–415.

Abate, J., Choudhury, G. L., & Whitt, W. 2000. *An Introduction to Numerical Transform Inversion and Its Application to Probability Models*. Computational Probability—Springer, 257–323.

Abate, J., & Whitt, W. 1992. Numerical inversion of probability generating functions. *Operations Research Letters*, **12**(4), 245–251.

Abramowitz, M., & Stegun, I. A. 1965. *Handbook of Mathematical Functions*. New York: Dover Publications.

Abramowitz, M., & Stegun, I. A. 1972. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards Applied Mathematics Series 55. Tenth Printing. ERIC.

Acerbi, C. 2002. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking & Finance*, **26**(7), 1505–1518.

Acerbi, C. 2004. Coherent representations of subjective risk-aversion. In: Szego, G. (ed.), *Risk Measures for the 21st Century*. New York: Wiley, 147–207.

Ahmed, N., Natarajan, T., & Rao, K. R. 1974. Discrete cosine transform. *IEEE Transactions on Computers*, **100**(1), 90–93.

Aigner, D. J., Amemiya, T., & Poirier, D. J. 1976. On the estimation of production frontiers: maximum likelihood estimation of the parameters of a discontinuous density function. *International Economic Review*, **17**(2), 377–396.

Akgiray, V., & Lamoureux, C. G. 1989. Estimation of stable-law parameters: A comparative study. *Journal of Business & Economic Statistics*, **7**(1), 85–93.

Albrecher, H., Asmussen, S., & Kortschak, D. 2006. Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes*, **9**(2), 107–130.

Albrecher, H., Hipp, C., & Kortschak, D. 2010. Higher-order expansions for compound distributions and ruin probabilities with subexponential claims. *Scandinavian Actuarial Journal*, **2010**(2), 105–135.

Alder, R., Feldman, R., & Taqqu, M. S. 1998. *A Practical Guide to Heavy-Tails: Statistical Techniques for Analysing Heavy-Tailed Distributions*. Boston, MA: Birkhäuser.

Aldrovandi, R. 2001. *Special Matrices of Mathematical Physics: Stochastic, Circulant, and Bell Matrices*. World Scientific.

Andrews, D. F., & Mallows, C. L. 1974. Scale mixtures of normal distributions. *Journal of the Royal Statistical Society, Series B (Methodological)*, **36**(1), 99–102.

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.

© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

Andrews, Larry C 1985. *Special Functions for Engineers and Applied Mathematicians*. Macmillan New York.

Antoniadis, A., Feuerverger, A., & Gonçalves, P. 2006. Wavelet-based estimation for univariate stable laws. *Annals of the Institute of Statistical Mathematics*, **58**(4), 779–807.

Arnold, B. C. 1973. Some characterizations of the exponential distribution by geometric compounding. *SIAM Journal on Applied Mathematics*, **24**(2), 242–244.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. 1997. Thinking coherently: generalised scenarios rather than VAR should be used when calculating regulatory capital. *RISK Magazine*, **10**(11), 68–71.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. 1999. Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Asheim, A., Deano, A., Huybrechs, D., & Wang, H. 2012. *A Gaussian quadrature rule for oscillatory integrals on a bounded interval. arXiv preprint arXiv:1212.1293.*

Asmussen, S., & Rojas-Nandayapa, L. 2008. Asymptotics of sums of lognormal random variables with Gaussian copula. *Statistics & Probability Letters*, **78**(16), 2709–2714.

Azzalini, A. 1985. A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**(2), 171–178.

von Bahr, B. 1975. Asymptotic ruin probabilities when exponential moments do not exist. *Scandinavian Actuarial Journal*, **1975**(1), 6–10.

Baker, C. T. H. 2000. A perspective on the numerical treatment of Volterra equations. *Journal of computational and applied mathematics*, **125**(1), 217–249.

Balanda, K. P., & MacGillivray, H. L. 1988. Kurtosis: a critical review. *American Statistician*, **42**(2), 111–119.

Balanda, K. P., & MacGillivray, H. L. 1990. Kurtosis and spread. *Canadian Journal of Statistics*, **18**(1), 17–30.

Baleanu, D., & Trujillo, J. I. 2010. A new method of finding the fractional Euler–Lagrange and Hamilton equations within Caputo fractional derivatives. *Communications in Nonlinear Science and Numerical Simulation*, **15**(5), 1111–1115.

Balkema, G., Balkema, A. A., & Embrechts, P. 2007. *High Risk Scenarios and Extremes: A Geometric Approach*. European Mathematical Society.

Balkema, A. A., & De Haan, L. 1972. On R. Von Mises' Condition for the Domain of Attraction of $\exp(-e^{-x})$. *Annals of Mathematical Statistics*, **43**(4), 1352–1354.

Balkema, A. A., Geluk, J. L., & De Haan, L. 1979. An extension of Karamata's Tauberian theorem and its connection with complementary convex functions. *Quarterly Journal of Mathematics*, **30**(4), 385–416.

Baltrūnas, A. 1999. Second order behaviour of ruin probabilities. *Scandinavian Actuarial Journal*, **1999**(2), 120–133.

Baltrūnas, A., & Omey, E. 1998. The rate of convergence for subexponential distributions. *Lithuanian Mathematical Journal*, **38**(1), 1–14.

Baltrūnas, A., Omey, E., & Van Gulck, S. 2006. Hazard rates and subexponential distributions. *Publications de l'Institut Mathématique. Nouvelle Série*, **80**(94), 29–46.

Barbe, P., & McCormick, W. P. 2004. *Asymptotic expansions for infinite weighted convolutions of heavy tail distributions and applications. arXiv preprint math/0412537.*

Barbe, P., & McCormick, W. P. 2005a. Asymptotic expansions of convolutions of regularly varying distributions. *Journal of the Australian Mathematical Society*, **78**(3), 339–371.

Barbe, P., & McCormick, W. P. 2005b. *Asympyotic expansions for infinite weighted convolutions of light subexponential distributions. arXiv preprint math/0512141.*

Barbe, P., & McCormick, W. P. 2009a. *Asymptotic Expansions for Infinite Weighted Convolutions of Heavy Tail Distributions and Applications*. AMS Bookstore.

Barbe, P., & McCormick, W. P. 2009b. *Asymptotic Expansions for Infinite Weighted Convolutions of Heavy Tail Distributions and Applications*. Memories of the American Mathematical Society, **922**.

Barbe, P., McCormick, W. P., & Zhang, C. 2007. Asymptotic expansions for distributions of compound sums of random variables with rapidly varying subexponential distribution. *Journal of Applied Probability*, **44**(3), 670–684.

Bardoutsos, A. G., & Konstantinides, D. G. 2011. Characterization of tails through hazard rate and convolution closure properties. *Journal of Applied Probability*, **48A** (2011), 123–132.

Barndorff-Nielsen, O. E. 1997a. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, **24**(1), 1–13.

Barndorff-Nielsen, O. E. 1997b. Processes of normal inverse Gaussian type. *Finance and Stochastics*, **2**(1), 41–68.

Barndorff-Nielsen, O. E., & Shephard, N. 2001. *Normal modified stable processes*. Discussion Paper Series 72. Department of Economics, University of Oxford.

Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. 1998. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, **85**(3), 549–559.

Baxter, G. 1961. An analytic approach to finite fluctuation problems in probability. *Journal d'Analyse Mathématique*, **9**(1), 31–70.

BCBS. 2002. Quantitative impact study for operational risk: overview of individual loss data and lessons learned. Basel Committee on Banking Supervision, Bank for International Settlements.

BCBS. 2004. International convergence of capital measurement and capital standards. Basel Committee on Banking Supervision, Bank for International Settlements.

BCBS. 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework (comprehensive version). Basel Committee on Banking Supervision, Bank for International Settlements.

Beaumont, M. A., Cornuet, J-M., Marin, J.-M., & Robert, C. P. 2009. Adaptive approximate Bayesian computation. *Biometrika*, **96**(4), 983–990.

Beaumont, M. A., Zhang, W., & Balding, D. J. 2002. Approximate Bayesian computation in population genetics. *Genetics*, **162**(4), 2025–2035.

Beirlant, J., Dierckx, G., Goegebeur, Y., & Matthys, G. 1999. Tail index estimation and an exponential regression model. *Extremes*, **2**(2), 177–200.

Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., Waal, D. D., & Ferro, C. 2004. *Statistics of Extremes: Theory and Applications*. Wiley.

Beirlant, J., Teugels, J. L., & Vynckier, P. 1996. *Practical Analysis of Extreme Values*. Leuven University Press.

Belov, I. A. 2005. On the computation of the probability density function of α-stable distributions. *Mathematical Modelling and Analysis, Proceedings of the 10th International Conference MMA2005&CMAM2, Trakai, Technika*, **2005**, 333–341.

Benktander, G., & Segerdahl, C. O. 1960. On the analytical representation of claim distributions with special reference to excess of loss reinsurance. *Transactions 16-th International Congress Actuaries*, 626–646.

Bentkus, V., Götze, F., & Paulauskas, V. 1996. Bounds for the accuracy of Poissonian approximations of stable laws. *Stochastic Processes and Their Applications*, **65**(1), 55–68.

Bentkus, V., Juozulynas, A., & Paulauskas, V. 2001. Lévy–LePage series representation of stable vectors: convergence in variation. *Journal of Theoretical Probability*, **14**(4), 949–978.

Bergstrom, H. 1953. On some expansions of stable distributional functions. *Arkiv för Matematik*, **2**(4), 375–378.

Berry, A. C. 1941. The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, **49**(1), 122–139.

Besbeas, P., & Morgan, B. J. T. 2008. Improved estimation of the stable laws. *Statistics and Computing*, **18**(2), 219–231.

Biagini, F. U., & Ulmer, S. 2009. Asymptotics for operational risk quantified with expected shortfall. *ASTIN Bulletin*, **39**(2), 735.

Bianchi, M. L., Rachev, S. T., Kim, Y. S., & Fabozzi, F. J. 2010. *Tempered stable distributions and processes in finance: numerical analysis*, Mathematical and Statistical Methods for Actuarial Sciences and Finance. Springer, 33–42.

Bianchi, M. L., Rachev, S. T., Kim, Y. S., & Fabozzi, F. J. 2011. Tempered infinitely divisible distributions and processes. *Theory of Probability & its Applications*, **55**(1), 2–26.

Billingsley, P. 1995. *Probability and Measure*. John Wiley & Sons.

Bingham, N. H., Goldie, C. M., & Teugels, J. L. 1989. *Regular Variation*. Vol. **27**. Cambridge University Press.

Blattberg, R., & Sargent, T. 1971. Regression with non-Gaussian stable disturbances: some sampling results. *Econometrica*, **39**(3), 501–510.

Block, H. W., Savits, T. H., & Shaked, M. 1982. Some concepts of negative dependence. *Annals of Probability*, **10**(3), 765–772.

Blundell-Wignall, A., & Atkinson, P. 2010. Thinking Beyond Basel III: necessary solutions for capital and liquidity. *OECD Journal: Financial Market Trends*, **2010**(1), 5–6.

Böcker, K., & Klüppelberg, C. 2005. Operational VaR: a closed-form approximation. *Risk Magazine*, **12**, 90–93.

Böcker, K., & Klüppelberg, C. 2008. Modeling and measuring multivariate operational risk with Lévy copulas. *Journal of Operational Risk*, **3**(2), 3–27.

Böcker, K., & Sprittulla, J. 2006. Operational VAR: meaningful means. *Risk Magazine*, **12**, 96–98.

Bohman, H. 1975. Numerical inversions of characteristic functions. *Scandinavian Actuarial Journal*, **1975**(2), 121–124.

Bojanic, R., & Karamata, J. 1963. On Slowly Varying Functions and Asymptotic Relations. Tech. Rept. DTIC Document.

Bojanic, R., & Seneta, E. 1971. Slowly varying functions and asymptotic relations. *Journal of Mathematical Analysis and Applications*, **34**(2), 302–315.

Bondesson, L. 2006. Generalized gamma convolutions, Encyclopedia of Statistical Sciences, 4, John Wiley & Sons, Ltd.

Bondesson, L., Kristiansen, G. K., & Steutel, F. W. 1996. Infinite divisibility of random variables and their integer parts. *Statistics & Probability Letters*, **28**(3), 271–278.

Borovkov, A. A., & Borovkov, K. A. 2002. On probabilities of large deviations for random walks. I. Regularly varying distribution tails. *Theory of Probability & its Applications*, **46**(2), 193–213.

Borovkov, A. A., & Borovkov, K. A. 2010. *Asymptotic Analysis of Random Walks: Heavy-Tailed Distributions*. vol. **118**. Cambridge University Pres.

Boyarchenko, S. I., & Levendorskii, S. Z. 2000. Option pricing for truncated Lévy processes. *International Journal of Theoretical and Applied Finance*, **3**(03), 549–552.

Boyd, J. P. 2001. *Chebyshev and Fourier Spectral Methods*. Courier Dover Publications.

Brazauskas, V., Jones, B. L., Puri, M. L., & Zitikis, R. 2008. Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference*, **138**(11), 3590–3604.

Brix, A. 1999. Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, **31**(4), 929–953.

Buckle, D. J. 1995. Bayesian inference for stable distributions. *Journal of the American Statistical Association*, **90**(430), 605–613.

Bürmann, H. H. 1796. Essai de calcul fonctionnaire aux constantes ad-libitum. In: *Versuch einer vereinfachten Analysis; ein Auszug eines Auszuges von Herrn Bürmann* , vol. **2**, 495–499.

Burr, I. W. 1942. Cumulative frequency functions. *Annals of Mathematical Statistics*, **13**(2), 215–232.

Burr, I. W., & Cislak, P. J. 1968. On a general system of distributions: I. Its curve-shape characteristics; II. The sample median. *Journal of the American Statistical Association*, **63**(322), 627–635.

Butzer, P. L., & Hahn, L. 1978. General theorems on rates of convergence in distribution of random variables II. Applications to the stable limit laws and weak law of large numbers. *Journal of Multivariate Analysis*, **8**(2), 202–221.

Calvetti, D., Golub, G., Gragg, W., & Reichel, L. 2000. Computation of Gauss-Kronrod quadrature rules. *Mathematics of Computation of the American Mathematical Society*, **69**(231), 1035–1052.

Candan, C., Kutay, M. A., & Ozaktas, H. M. 2000. The discrete fractional Fourier transform. *IEEE Transactions on Signal Processing*, **48**(5), 1329–1337.

Cappé, O., Guillin, A., Marin, J. M., & Robert, C. P. 2004. Population monte carlo. *Journal of Computational and Graphical Statistics*, **13**(4), 907–929.

Carr, P., Geman, H., Madan, D. B., & Yor, M. 2002. The fine structure of asset returns: an empirical investigation. *Journal of Business*, **75**(2), 305–333.

Casarin, R. 2003. Bayesian inference for mixtures of stable distributions. Working paper No. 0428, CEREMADE, University Paris IX.

del Castillo, J., & Daoudi, J. 2009. Estimation of the generalized Pareto distribution. *Statistics & Probability Letters*, **79**(5), 684–688.

Castillo, E., & Hadi, A. S. 1997. Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, **92**(440), 1609–1620.

Chambers, J. M., Mallows, C. L., & Stuck, B. W. 1976. A method for simulating stable random variables. *Journal of the American Statistical Association*, **71**(354), 340–344.

Chan, T. 1999. Pricing contingent claims on stocks driven by Lévy processes. *Annals of Applied Probability*, **9**(2), 504–528.

Chavez-Demoulin, V., Embrechts, P., & Nešlehová, J. 2006. Quantitative models for operational risk: extremes, dependence and aggregation. *Journal of Banking & Finance*, **30**(10), 2635–2658.

Chen, Y., & Zhang, W. 2007. Large deviations for random sums of negatively dependent random variables with consistently varying tails. *Statistics & Probability Letters*, **77**(5), 530–538.

Cheng, R. C. H., & Liu, W. B. 1997. A continuous representation of the family of stable law distributions. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **59**(1), 137–145.

Cheng, S., & Peng, L. 2001. Confidence intervals for the tail index. *Bernoulli*, **7**(5), 751–760.

Chernin, K. E., & Ibragimov, I. A. 1959. On the unimodality of stable laws. *Teor. Veroyatnost. i Primenen*, **4**, 453–456.

Chernobai, A., & Rachev, S. T. 2004. Stable modelling of operational risk. In: Cruz, M. G. (ed.), *Operational Risk Modelling and Analysis: Theory and Practice*. London: Risk Books.

Chernoff, H., Gastwirth, J. L., & Johns, M. V. 1967. Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Annals of Mathematical Statistics*, **38**(1), 52–72.

Chhikara, R. S., & Folks, J. L. 1989. *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*. Vol. **95**. CRC Press.

Chistyakov, V. P. 1964. A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory of Probability & its Applications*, **9**(4), 640–648.

Chover, J, Ney, P, & Wainger, S. 1973. Degeneracy properties of subcritical branching processes. *Annals of Probability*, **1**(4), 663–673.

Christoph, G., & Schreiber, K. 1998a. Discrete stable random variables. *Statistics & Probability Letters*, **37**(3), 243–247.

Christoph, G., & Schreiber, K. 1998b. *The Generalized Discrete Linnik Distributions*. Springer.

Clenshaw, C. W., & Curtis, A. R. 1960. A method for numerical integration on an automatic computer. *Numerische Mathematik*, **2**(1), 197–205.

Cline, D. B. H. 1994. Intermediate regular and Π variation. *Proceedings of the London Mathematical Society*, **3**(3), 594–616.

Cline, D. B. H., & Samorodnitsky, G. 1994. Subexponentiality of the product of independent random variables. *Stochastic Processes and Their Applications*, **49**(1), 75–98.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer Verlag.

Coles, S., Heffernan, J., & Tawn, J. 1999. Dependence measures for extreme value analyses. *Extremes*, **2**(4), 339–365.

Coles, S. G., & Powell, E. A. 1996. Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review/Revue Internationale de Statistique*, **64**(1), 119–136.

Coles, S. G., & Tawn, J. A. 1991. Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B (Methodological)*, **53**(2), 377–392.

Cont, R., Potters, M., & Bouchaud, J.-P. 1997. *Scaling in stock market data: stable laws and beyond. arXiv preprint cond-mat/9705087.*

Cont, R., & Tankov, P. 2004. *Financial Modelling With Jump Processes*. Boca Raton, FL: Chapman & Hall/CRC.

Cope, E. W., Antonini, G., Mignola, G., & Ugoccioni, R. 2009. Challenges and pitfalls in measuring operational risk from loss data. *Journal of Operational Risk*, **4**(4), 3–27.

Cox, D. R., & Isham, V. 1980. *Point Processes*. Vol. **12**. Chapman & Hall/CRC.

Cruz, M. G. 2002. *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons New York.

Cruz, M. G. (ed.) 2004. *Operational Risk Modelling and Analysis: Theory and Practice*. London: Risk Books.

Cruz, M. G., Peters, G. W., & Shevchenko, P. V. 2015. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. John Wiley & Sons.

Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. 2010. Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution*, **25**(7), 410–418.

Daley, D. J., Omey, E., & Vesilo, R. 2007. The tail behaviour of a random sum of subexponential random variables and vectors. *Extremes*, **10**(1), 21–39.

Daniels, H. E. 1954. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, **25**(4), 631–650.

Daniels, H. E. 1987. Tail probability approximations. *International Statistical Review/Revue Internationale de Statistique*, **55**(1), 37–48.

Daniélsson, J., Embrechts, *P.*, Goodhart, C., Keating, C., Muennich, F., Renault, O., & Shin, H. S. 2001. *An academic response to Basel II. Special Paper No. 130, LSE Financial Markets Group*.

Daniélsson, J., De Haan, L., Peng, L., & De Vries, C. G. 2001. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis*, **76**(2), 226–248.

Daniélsson, J., & De Vries, C. G. 1997. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, **4**(2), 241–257.

Daniélsson, J., Hartmann, P., & De Vries, C. 1998. The cost of conservatism. *RISK,* **11**(1), 101–103.

David, H. A. 1968. Miscellanea: Gini's mean difference rediscovered. *Biometrika*, **55**(3), 573–575.

David, H. A., & Nagaraja, H. N. 1970. *Order Statistics*. Wiley Online Library.

Davies, R. B. 1973. Numerical inversion of a characteristic function. *Biometrika*, **60**(2), 415–417.

Davison, A. C. 1984. Modelling excesses over high thresholds, with an application. In: de Oliveira, J. T. (ed.), *Statistical Extremes and Applications*, 461–482. Springer, Netherlands.

De Bruijn, N. G. 1970. *Asymptotic Methods in Analysis*. Vol. **4**. DoverPublications. com.

De Haan, L., & Resnick, S. 1996. Second-order regular variation and rates of convergence in extreme-value theory. *Annals of Probability*, **24**(1), 97–124.

De Luca, G., & Rivieccio, G. 2012. Multivariate tail dependence coefficients for Archimedean Copulae. In: Di Ciaccio, A., Coli, M., Angulo, I., & Jose, M. (eds), *Advanced Statistical Methods for the Analysis of Large Data-Sets*. Springer, 287–296.

De Pril, N. 1986. On the exact computation of the aggregate claims distribution in the individual life model. *ASTIN Bulletin*, **16**(2), 109–112.

De Pril, N. 1988. Improved approximations for the aggregate claims distribution of a life insurance portfolio. *Scandinavian Actuarial Journal*, **1988**(1-3), 61–68.

De Pril, N. 1989. The aggregate claims distribution in the individual model with arbitrary positive claims. *ASTIN Bulletin*, **19**(1), 9–24.

De Smit, J. H. A. 1973. A simple analytic proof of the Pollaczek-Wendel identity for ordered partial sums. *Annals of Probability*, **1**(2), 348–351.

De Vylder, F., & Goovaerts, M. J. 1994. A note on the solution of practical ruin problems. *Insurance: Mathematics and Economics*, **15**(2), 181–186.

Decamps, J.-P., Rochet, J.-C., & Roger, B. 2004. The three pillars of Basel II: optimizing the mix. *Journal of Financial Intermediation*, **13**(2), 132–155.

Degen, M. 2010. The calculation of minimum regulatory capital using single-loss approximations. *Journal of Operational Risk*, **5**(4), 1–15.

Degen, M., & Embrechts, P. 2011. Scaling of high-quantile estimators. *Journal of Applied Probability*, **48**(4), 968–983.

Degen, M., Embrechts, P., & Lambrigger, D. D. 2007. The quantitative modeling of operational risk: between g-and-h and EVT. *ASTIN Bulletin*, **37**(2), 265–291.

Dekkers, A. L. M., & De Haan, L. 1989. On the estimation of the extreme-value index and large quantile estimation. *Annals of Statistics*, **17**(4), 1795–1832.

Del Moral, P. 2004. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer.

Del Moral, P., Doucet, A., & Jasra, A. 2006. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **68**(3), 411–436.

Del Moral, P., Doucet, A., & Jasra, A. 2012. An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, **22**(5), 1009–1020.

Del Moral, P., Peters, G. W., & Vergé, C. 2013. *An introduction to particle integration methods: with applications to risk and insurance*. In: Dick, J., Kuo, F. Y., Peters, G. W., & Sloan, I. H. (eds), *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer.

Denuit, M., Genest, C., & Marceau, É. 1999. Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, **25**(1), 85–104.

Devroye, L. 1981. On the computer generation of random variables with a given characteristic function. *Computers & Mathematics with Applications*, **7**(6), 547–552.

Devroye, L. 1984. Methods for generating random variates with Polya characteristic functions. *Statistics & Probability Letters*, **2**(5), 257–261.

Devroye, L. 1993. A triptych of discrete distributions related to the stable law. *Statistics & Probability Letters*, **18**(5), 349–351.

Devroye, L. 2009. Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **19**(4), 18.

Devroye, L., & Devroye, L. 1986. *Non-Uniform Random Variate Generation*. Vol. **4**. Springer-Verlag New York.

Dhaene, J., & Vandebroek, M. 1995. Recursions for the individual model. *Insurance: Mathematics and Economics*, **16**(1), 31–38.

Dickson, D. C. M. 1995. A review of Panjer's recursion formula and its applications. *British Actuarial Journal*, **1**(1), 107–124.

Diebold, F. X., Schuermann, T., & Stroughair, J. D. 2000. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, **1**(2), 30–35.

Doeblin, W. 1940. Sur l'ensemble de puissances d'une loi de probabilité. *Studia mathematica*, **9**(1), 71–96.

Doganoglu, T., & Mittnik, S. 1998. An approximation procedure for asymmetric stable paretian densities. *Computational Statistics*, **13**(4), 463–476.

Doucet, A., Godsill, S., & Andrieu, C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, **10**(3), 197–208.

Doucet, A., & Johansen, A. M. 2009. A tutorial on particle filtering and smoothing: fifteen years later. *Handbook of Nonlinear Filtering*, Oxford University Press, **12**, 656–704.

Doucet, A., Johansen, A. M., & Tadić, V. B. 2010. On solving integral equations using Markov Chain Monte Carlo methods. *Applied Mathematics and Computation*, **216**(10), 2869–2880.

Dowd, K., & Blake, D. 2006. After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *Journal of Risk and Insurance*, **73**(2), 193–229.

Downton, F. 1966. Linear estimates with polynomial coefficients. *Biometrika*, **53**(1/2), 129–141.

Dubey, S. D. 1970. Compound gamma, beta and F distributions. *Metrika*, **16**(1), 27–31.

DuMouchel, W. H. 1971. *Stable Distributions in Statistical Inference*. University Microfilms.

DuMouchel, W. H. 1975. Stable distributions in statistical inference: 2. Information from stably distributed samples. *Journal of the American Statistical Association*, **70**(350), 386–393.

DuMouchel, W. H. 1983. Estimating the stable index α in order to measure tail thickness: a critique. *Annals of Statistics*, **11**(4), 1019–1031.

Dunn, P. K., & Smyth, G. K. 2005. Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, **15**(4), 267–280.

Dupuis, D. J. 1999. Exceedances over high thresholds: a guide to threshold selection. *Extremes*, **1**(3), 251–261.

Durrett, R. 2010. *Probability: Theory and Examples*. Vol. **3**. Cambridge University Press.

Dutta, K. K., & Babbel, D. F. 2002. *On measuring skewness and kurtosis in short rate distributions: the case of the us dollar london inter bank offer rates.* Center for Financial Institutions Working Papers 02, 25.

Dutta, K., & Perry, J. 2006. *A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital*. Federal Reserve Bank of Boston. Working paper No. 06-13.

Efron, B. F., & Tibshirani, R. J. 1993. *An Introduction to the Bootstrap*. London: Chapman and Hall.

Egorov, V. A. 1980. On the rate of convergence to a stable law. *Theory of Probability and its Applications*, **25**(1), 180–187.

Elamir, E. A. H., & Seheult, A. H. 2003. Trimmed L-moments. *Computational Statistics & Data Analysis*, **43**(3), 299–314.

Embrechts, P., & Goldie, C. M. 1982. On convolution tails. *Stochastic Processes and their Applications*, **13**(3), 263–278.

Embrechts, P., Grübel, R., & Pitts, S. M. 1993. Some applications of the fast Fourier transform algorithm in insurance mathematics This paper is dedicated to Professor WS Jewell on the occasion of his 60th birthday. *Statistica Neerlandica*, **47**(1), 59–75.

Embrechts, P., & Hofert, M. 2010. A note on generalized inverses. *Mathematical Methods of Operations Research*, **77**(3), 423–432.

Embrechts, P., Klüppelberg, C., & Mikosch, T. 1997. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

Embrechts, P., Maejima, M., & Teugels, J. 1985a. Asymptotic behaviour of compound distributions. *ASTIN Bulletin*, **15**(1), 45–48.

Embrechts, P., Maejima, M., & Omey, E. 1985b. Some limit theorems for generalized renewal measures. *Journal of the London Mathematical Society*, **2**(1), 184–192.

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. 1999. Extreme value theory as a risk management tool. *North American Actuarial Journal*, **3**(2), 30–41.

Embrechts, P., & Schmidli, H. 1994. Modelling of extremal events in insurance and finance. *Zeitschrift für Operations Research*, **39**(1), 1–34.

Embrechts, P., & Veraverbeke, N. 1982. Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics*, **1**(1), 55–72.

Eriksson, A., Ghysels, E., & Wang, F. 2009. The normal inverse Gaussian distribution and the pricing of derivatives. *Journal of Derivatives*, **16**(3), 23–37.

Eugenia, C. M., & Cabras, S. 2007. A default Bayesian procedure for the generalized Pareto distribution. *Journal of Statistical Planning and Inference*, **137**(2), 473–483.

Evans, G. A., & Webster, J. R. 1999. A comparison of some methods for the evaluation of highly oscillatory integrals. *Journal of Computational and Applied Mathematics*, **112**(1), 55–69.

Fama, E. 1965. The behaviour of stock market prices. *Journal of Business*, **38**(1), 34–105.

Fama, E. F., & Roll, R. 1968. Some properties of symmetric stable distributions. *Journal of the American Statistical Association*, **63**(323), 817–836.

Faÿ, G., González-Arévalo, B., Mikosch, T., & Samorodnitsky, G. 2006. Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Systems*, **54**(2), 121–140.

Feldman, Raya, & Taqqu, Murad. 1998. *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Springer.

Feller, W. 1945. The fundamental limit theorems in probability. *Bulletin of the American Mathematical Society*, **51**(11), 800–832.

Feller, W. 1966. *An Introduction to Probability Theory*. Vol. **II**. New York: John Wiley & Sons.

Feller, W. 2008. *An Introduction to Probability Theory and its Applications*. Vol. **2**. John Wiley & Sons.

Feng, J., Li, J., Gao, L., & Hua, Z. 2012. A combination model for operational risk estimation in a Chinese banking industry case. *Journal of Operational Risk*, **7**(2), 17.

Feuerverger, A., & Hall, P. 1999. Estimating a tail exponent by modelling departure from a Pareto distribution. *Annals of Statistics*, **27**(2), 760–781.

Feuerverger, A., & McDunnough, P. 1981. On the efficiency of empirical characteristic function procedures. *Journal of the Royal Statistical Society, Series B (Methodological)*, **43**(1), 20–27.

Field, Christopher, & Genton, Marc, G. 2006. The multivariate g-and-h distribution. *Technometrics*, **48**(1), 104–111.

Fischer, M. 2010. Generalized Tukey-type distributions with application to financial and teletraffic data. *Statistical Papers*, **51**(1), 41–56.

Fischer, M., Horn, A., & Klein, I. 2007. Tukey-type distributions in the context of financial data. *Communications in Statistics Theory and Methods*, **36**(1), 23–35.

Fischer, M., & Klein, I. 2004. Kurtosis modelling by means of the J-transformation. *Allgemeines Statistisches Archiv*, **88**(1), 35–50.

Fisher, R. A., & Tippett, L. H. C. 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. **24**(2). Cambridge Univ Press, 180–190.

Fofack, H., & Nolan, J. P. 1999. Tail behavior, modes and other characteristics of stable distributions. *Extremes*, **2**(1), 39–58.

Folks, J. L., & Chhikara, R. S. 1978. The inverse Gaussian distribution and its statistical application–a review. *Journal of the Royal Statistical Society, Series B (Methodological)*, **40**(3), 263–289.

Foss, S., Korshunov, D., & Zachary, S. 2011. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Vol. **38**. Springer.

Foss, S., Korshunov, D., & Zachary, S. 2013. Introduction. In: *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer, 1–6.

Foss, S., & Richards, A. 2010. On sums of conditionally independent subexponential random variables. *Mathematics of Operations Research*, **35**(1), 102–119.

Fox, C. 1961. The G and H functions as symmetrical Fourier kernels. *Transactions of the American Mathematical Society*, **98**(1961), 395–429.

Frachot, A., Moudoulaud, O., & Roncalli, T. 2004. Loss distribution approach in practice. In: Ong, M. (ed.), *The Basel Handbook: A Guide for Financial Practitioners*. Risk Books.

Frahm, G., Junker, M., & Schmidt, R. 2005. Estimating the tail-dependence coefficient: Properties and pitfalls. *Insurance: Mathematics and Economics*, **37**(1), 80–100.

Frank, M. J., Nelsen, R. B., & Schweizer, B. 1987. Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields*, **74**(2), 199–211.

Garcia, R., Renault, E., & Veredas, D. 2006. *Estimation of Stable Distributions by Indirect Inference*. CORE.

Garoni, T. M., & Frankel, N. E. 2002. Lévy flights: exact results and asymptotics beyond all orders. *Journal of Mathematical Physics*, **43**(5), 2670–2689.

Gawronski, W., & Wiessner, M. 1992. Asymptotics and inequalities for the mode of stable laws. *Statistics and Risk Modelling*, **10**(1-2), 183–198.

Geluk, J. L. 1981. n-Regular Variation. *Proceedings of the American Mathematical Society*, **82**(4), 565–570.

Geluk, J. L. 1992. Second order tail behaviour of a subordinated probability distribution. *Stochastic Processes and their Applications*, **40**(2), 325–337.

Geluk, J., de Haan, L., Resnick, S., & Stărică, C. 1997. Second-order regular variation, convolution and the central limit theorem. *Stochastic Processes and their Applications*, **69**(2), 139–159.

Geluk, J., & Ng, K. W. 2006. Tail behavior of negatively associated heavy-tailed sums. *Journal of Applied Probability*, **43**(2), 587–593.

Geluk, J. L., Peng, L., & De Vries, C. G. 2000. Convolutions of heavy-tailed random variables and applications to portfolio diversification and MA (1) time series. *Advances in Applied Probability*, **32**(4), 1011–1026.

Geluk, J., & Tang, Q. 2009. Asymptotic tail probabilities of sums of dependent subexponential random variables. *Journal of Theoretical Probability*, **22**(4), 871–882.

Gerber, H. U. 1982. On the numerical evaluation of the distribution of aggregate claims and its stop-loss premiums. *Insurance: Mathematics and Economics*, **1**(1), 13–18.

Gerber, H. U., & Jones, D. 1976. Some practical considerations in connection with the calculation of stop-loss premiums. *Transactions of the Society of Actuaries*, **28**, 215–31.

Gerhold, S., Schmock, U., & Warnung, R. 2010. A generalization of Panjers recursion and numerically stable risk aggregation. *Finance and Stochastics*, **14**(1), 81–128.

Ghosh, M. 1981. Multivariate negative dependence. *Communications in Statistics-Theory and Methods*, **10**(4), 307–337.

Gil-Pelaez, J. 1951. Note on the inversion theorem. *Biometrika*, **38**(3-4), 481–482.

Gilks, W. R., & Berzuini, C. 2002. Following a moving target Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **63**(1), 127–146.

Givens, G. H., & Raftery, A. E. 1996. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association*, **91**(433), 132–141.

Gnedenko, B. 1939. the theory of limiting theorems for sums of independent random variables. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya*, **3**(2), 181–232.

Gnedenko, B. V. 1943. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44**(3), 423–453.

Gnedenko, B. V. 1983. On some stability theorems. In: *Stability Problems for Stochastic Models*. Springer, 24–31.

Gnedenko, B. V., & Kolmogorov, A. N. 1968. *Limit Distributions for Sums of Independent Random Variables*. Addison-Wesley.

Godsill, S. 1999. MCMC and EM-based methods for inference in heavy-tailed processes with alpha stable innovations. In: *Proceedings IEEE Signal Processing Workshop on Higher Order Statistics*, 228–232.

Godsill, S. 2000a. Inference in symmetric alpha-stable noise using MCMC and the slice sampler. In: *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. **VI**, 3806–3809.

Godsill, S. 2000b. Inference in symmetric alpha-stable noise using MCMC and the slice sampler. *Acoustics, Speech, and Signal Processing, 2000*, ICASSP'00. Proceedings. 2000 on IEEE International Conference, Vol. **6**. IEEE, 3806–3809.

Goldie, C. M. 1978. Subexponential distributions and dominated-variation tails. *Journal of Applied Probability*, **15**, 440–442.

Gomes, M. I., & De Haan, L. 1999. Approximation by penultimate extreme value distributions. *Extremes*, **2**(1), 71–85.

Goovaerts, M., & De Vylder, F. 1984. A stable recursive algorithm for evaluation of ultimate ruin probabilities. *ASTIN Bulletin*, **14**(1), 53–60.

Greenwood, P. 1973. Asymptotics of randomly stopped sequences with independent increments. *Annals of Probability*, **1**(2), 317–321.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. 1979. Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, **15**(5), 1049–1054.

Gregoriou, G. N. 2009. *Operational Risk Toward Basel III: Best Practices and Issues in Modeling, Management, and Regulation*. Wiley.

Grimshaw, S. D. 1993. Computing maximum likelihood estimates for the generalized Pareto distribution. *Technometrics*, **35**(2), 185–191.

Grübel, R. 1987. On subordinated distributions and generalized renewal measures. *Annals of Probability*, **15**(1), 394–415.

Grübel, R., & Hermesmeier, R. 1999. Computation of compound distributions I: aliasing errors and exponential tilting. *ASTIN Bulletin*, **29**(2), 197–214.

Grübel, R., & Hermesmeier, R. 2000. Computation of compound distributions II: Discretization errors and Richardson extrapolation. *ASTIN Bulletin*, **30**(2), 309–331.

de Haan, L., & Ferreira, A. 2006. *Extreme Value Theory*. Springer.

Hall, P. 1981a. On the rate of convergence to a stable law. *Journal of the London Mathematical Society*, **2**(1), 179.

Hall, P. 1981b. Two-sided bounds on the rate of convergence to a stable law. *Probability Theory and Related Fields*, **57**(3), 349–364.

Hall, P. 1981c. A comedy of errors: the canonical form for a stable characteristic function. *Bulletin of the London Mathematical Society*, **13**(1), 23–27.

Hall, P. 1982. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B (Methodological)*, **44**(1), 37–42.

Hall, P. 1984. On unimodality and rates of convergence for stable laws. *Journal of the London Mathematical Society*, **2**(2), 371.

Hallinan, A. Jr. 1993. A review of the Weibull distribution. *Journal of Quality Technology*, **25**(2), 85–93.

Hardin, C. D. Jr. 1984. Skewed Stable Variables and Processes. Tech. Rept. DTIC Document.

He, X., & Fung, W. K. 1999. Method of medians for lifetime data with Weibull models. *Statistics in Medicine*, **18**(15), 1993–2009.

Headrick, T. C., Kowalchuk, R. K., & Sheng, Y. 2008. Parametric probability densities and distribution functions for Tukey g-and-h transformations and their use for fitting data. *Applied Mathematical Sciences*, **2**(9), 449–462.

Headrick, T. C., & Pant, M. D. 2012a. Characterizing Tukey h and hh-Distributions through L-Moments and L-Correlation. *ISRN Applied Mathematics*, **2012**, Article ID 980153, 20.

Headrick, T. C., & Pant, M. D. 2012b. A logistic L-moment-based analog for the Tukey g-h, g, h, and h-h system of distributions. *ISRN Probability and Statistics*, **2012**, Article ID 245986.

Hess, K. T., Liewald, A., & Schmidt, K. D. 2001. *An Extension of Panjer's Recursion*. Professoren des Inst. für Math. Stochastik.

Hill, B. M. 1975. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**(5), 1163–1174.

Hoaglin, D. C. 2006. Summarizing Shape Numerically: The g-and-h Distributions. In: Exploring Data Tables, Trends, and Shapes (eds D. C. Hoaglin, F. Mosteller and J. W. Tukey), Hoboken, NJ: John Wiley & Sons, Inc.

Hofert, M. 2011. Sampling exponentially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, **22**(1), 3.

Hoffmann-Jørgensen, J. 1993. Stable densities. *Theory of Probability & its Applications*, **38**(2), 350–355.

Holt, D. R., & Crow, E. L. 1973. Tables and graphs of the stable probability density functions. *Journal of Research of the National Bureau of Standards B*, **77**, 143–198.

Hosking, J. R. M. 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B (Methodological)*, **52**(1), 105–124.

Hosking, J. R. M., Wallis, J. R., & Wood, E. F. 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**(3), 251–261.

Hosking, J. R. M., & Wallis, J. R. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**(3), 339–349.

Houdré, C. 2002. Remarks on deviation inequalities for functions of infinitely divisible random vectors. *Annals of Probability*, **30**(3), 1223–1237.

Hougaard, P. 1986. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**(2), 387–396.

Hougaard, P., Lee, M.-L. T., & Whitmore, G. A. 1997. Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics*, **53**(4), 1225–1238.

Hu, C. Y., & Lin, G. D. 2003. Characterizations of the exponential distribution by stochastic ordering properties of the geometric compound. *Annals of the Institute of Statistical Mathematics*, **55**(3), 499–506.

Hua, L., & Joe, H. 2011. Second order regular variation and conditional tail expectation of multiple risks. *Insurance: Mathematics and Economics*, **49**(3), 537–546.

Huang, M. L., & Fung, K. Y. 1993. D-distribution and its applications. *Statistical Papers*, **34**(1), 143–159.

Hughett, P. 1998. Error bounds for numerical inversion of a probability characteristic function. *SIAM Journal on Numerical Analysis*, **35**(4), 1368–1392.

Humbert, P. 1945. Nouvelles correspondances symboliques. *Bulletin des Sciences Mathématiques*, **69**, 121–129.

Ibragimov, I. A. 1956. On the composition of unimodal distributions. *Theory of Probability & its Applications*, **1**(2), 255–260.

Ibragimov, I. A., & Linnik, V. 1971. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff.

Imai, J., & Kawai, R. 2011. On finite truncation of infinite shot noise series representation of tempered stable laws. *Physica A: Statistical Mechanics and its Applications*, **390**(23), 4411–4425.

Iserles, A., Nørsett, S. P., & Olver, S. 2006. Highly oscillatory quadrature: the story so far. In: *Numerical Mathematics and Advanced Applications* (eds. A.B. de Castro, D. Gómez, P. Quintela, P. Salgado). Proceedings of ENUMATH 2005, the 6th European Conference on Numerical Mathematics and Advanced Applications Santiago de Compostela, Spain, July 2005. Berlin: Springer, 97–118.

Jacobsen, M. 2006. *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Birkhauser.

Jacod, J., & Shiryaev, A. N. 1987. *Limit Theorems for Stochastic Processes*. Vol. **288**. Springer-Verlag Berlin.

Janicki, A., & Weron, A. 1994a. Can one see α-stable variables and processes? *Statistical Science*, **9**(1), 109–126.

Janicki, A., & Weron, A. 1994b. *Simulation and Chaotic Behavior of α-stable Stochastic Processes*. Vol. **178**. CRC Press.

Jasra, A., Doucet, A., Stephens, D. A., & Holmes, C. C. 2008. Interacting sequential Monte Carlo samplers for trans-dimensional simulation. *Computational Statistics & Data Analysis*, **52**(4), 1765–1791.

Jasra, A., Singh, S. S., Martin, J. S., & McCoy, E. 2012. Filtering via approximate Bayesian computation. *Statistics and Computing*, **22**(6), 1223–1237.

Jasra, A., Stephens, D. A., & Holmes, C. C. 2007. On population-based simulation for static inference. *Statistics and Computing*, **17**(3), 263–279.

Jelonek, P. 2012. Generating Tempered Stable Random Variates from Mixture Representation. Tech. Rept. University of Leicester, UK.

Jessen, A. H., & Mikosch, T. 2006. *Regularly Varying Functions*. University of Copenhagen, Laboratory of Actuarial Mathematics.

Jiménez, J. A., & Arunachalam, V. 2011. Using Tukey's g and h family of distributions to calculate value-at-risk and conditional value-at-risk. *Journal of Risk*, **13**(4), 95–116.

Joag-Dev, K., & Proschan, F. 1983. Negative association of random variables with applications. *Annals of Statistics*, **11**(1), 286–295.

Joe, H. 1990. Families of min-stable multivariate exponential and multivariate extreme value distributions. *Statistics & Probability Letters*, **9**(1), 75–81.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman&Hall.

Joe, H., Seshadri, V., & Arnold, B. C. 2012. Multivariate inverse gaussian and skew-normal densities. *Statistics & Probability Letters*, **82**(12), 2244–2251.

Johnson, N. L., Kotz, S., & Balakrishnan, N. 2002. *Continuous Multivariate Distributions: Models and Applications*. Vol. **1**. New York: John Wiley & Sons.

Jorge, M., & Boris, I. 1984. Some properties of the Tukey g and h family of distributions. *Communications in Statistics-Theory and Methods*, **13**(3), 353–369.

Jørgensen, B. 1986. Some properties of exponential dispersion models. *Scandinavian Journal of Statistics*, **13**(3), 187–197.

Jørgensen, B. 1987. Exponential dispersion models. *Journal of the Royal Statistical Society, Series B (Methodological)*, **49**(2), 127–162.

Jørgensen, B. 1991. *Exponential dispersion models*. VOl. **48**. Instituto de matematica pura e aplicada.

Jørgensen, B. 1997. *The Theory of Dispersion Models*. Vol. **76**. CRC Press.

Jørgensen, B., Martínez, J. R., & Tsao, M. 1994. Asymptotic behaviour of the variance function. *Scandinavian journal of statistics*, **21**(3), 223–243.

Jørgensen, B., Martínez, J. R., & Vinogradov, V. 2009. Domains of attraction to Tweedie distributions. *Lithuanian mathematical journal*, **49**(4), 399–425.

Jørgensen, B., & Paes De Souza, M. C. 1994. Fitting Tweedie's compound Poisson model to insurance claims data. *Scandinavian Actuarial Journal*, **1994**(1), 69–93.

Juárez, S. F., & Schucany, W. R. 2004. Robust and efficient estimation for the generalized pareto distribution. *Extremes*, **7**(3), 237–251.

Kagan, A. M., Linnik, I. U. V., Rao, C. R., & Ramachandran, B. 1973. *Characterization Problems in Mathematical Statistics*. New York: Wiley.

Kalemanova, A., Schmid, B., & Werner, R. 2007. The normal inverse Gaussian distribution for synthetic CDO pricing. *Journal of Derivatives*, **14**(3), 80–94.

Karl-Heinz, B. Y. 1994. On the exact calculation of the aggregate claims distribution in the individual life model, *ASTIN Bulletin*, **24**(1), 89–96.

Kashyap, A. K., & Stein, J. C. 2004. Cyclical implications of the Basel II capital standards. *Economic Perspectives-Federal Reserve Bank Of Chicago*, **28**(1), 18–33.

Katti, S. K. 1967. Infinite divisibility of integer-valued random variables. *Annals of Mathematical Statistics*, **38**(4), 1306–1308.

Kawai, R., & Masuda, H. 2009. *Exact simulation of finite variation tempered stable Ornstein-Uhlenbeck processes*. Preprint: Faculty of Mathematics, Kyushu University, Japan, kyushu-u.ac.jp, MI **2009-36**.

Kawai, R., & Masuda, H. 2011. On simulation of tempered stable random variates. *Journal of Computational and Applied Mathematics*, **235**(8), 2873–2887.

Khintchine, A. 1937. A new derivation of one formula by Lévy P. *Moscow State University Bulletin, I*, **1**, 1–5.

Khursheed, A., & Lai Saxena, K. M. 1981. Positive dependence in multivariate distributions. *Communications in Statistics-Theory and Methods*, **10**(12), 1183–1196.

Kienitz, J. 2010. *Tempered Stable Process*. Wiley Online Library.

Kilbas, A. A., Srivastava, H. M., & Trujillo, J. J. 2006. *Theory and Application of Fractional Differential Equations*. Elsevier.

Kim, K. J., Cools, R., & Ixaru, L. G. 2003. Extended quadrature rules for oscillatory integrands. *Applied Numerical Mathematics*, **46**(1), 59–73.

Kim, Y. S., Rachev, S. T., Bianchi, M. L., & Fabozzi, F. J. 2008. *A New Tempered Stable Distribution and its Application to Finance: Risk Assessment*. Springer, 77–109.

King, J. L. 2001. *Operational Risk: Measurements and Modelling*. John Wiley & Sons.

Kingman, J. F. C. 1992. *Poisson Processes*. Clarendon Press, Oxford Studies in Probability 3, Oxford University Press.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. 1998. *Loss Models: From Data to Decisions*. New York: John Wiley & Sons.

Klüppelberg, C. 1989. Subexponential distributions and characterizations of related classes. *Probability Theory and Related Fields*, **82**(2), 259–269.

Ko, B., & Tang, Q. 2008. Sums of dependent nonnegative random variables with subexponential tails. *Journal of Applied Probability*, **45**(1), 85–94.

Kokonendji, C. C., Dossou-Gbété, S., & Demétrio, C. G. B. 2004. Some discrete exponential dispersion models: poisson-Tweedie and Hinde-Demetrio classes. *SORT*, **28**(2), 201–214.

Komunjer, I. 2007. Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, **22**(5), 891–921.

Koponen, I. 1995. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E*, **52**(1), 1197.

Kornya, P. S. 2007. On approximating the individual risk model. *North American Actuarial Journal*, **11**(2),77–98.

Kortschak, D. 2012. Second order tail asymptotics for the sum of dependent, tail-independent regularly varying risks. *Extremes*, **15**(3), 353–388.

Kotz, S., & Nadarajah, S. 2000. *Extreme Value Distributions: Theory and Applications*. World Scientific Publishing Company.

Koutrouvelis, I. A. 1980. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, **75**(372), 918–928.

Koutrouvelis, I. A. 1981. An iterative procedure for the estimation of the parameters of stable laws: an iterative procedure for the estimation. *Communications in Statistics-Simulation and Computation*, **10**(1), 17–28.

Koutrouvelis, I. A. 1982. Estimation of location and scale in Cauchy distributions using the empirical characteristic function. *Biometrika*, **69**(1), 205–213.

Kozubowski, T. J., & Panorska, A. K. 1996. On moments and tail behavior of nu-stable random variables. *Statistics & Probability Letters*, **29**(4), 307–315.

Kozubowski, T. J., & Podgorski, K. 1999. A class of asymmetric distributions. *Actuarial Research Clearing House*, **1**, 113–134.

Kozubowski, T. J., & Podgórski, K. 2001. Asymmetric Laplace laws and modeling financial data. *Mathematical and Computer Modelling*, **34**(9), 1003–1021.

Küchler, U., & Sorensen, M. 1997. *Exponential Families of Stochastic Processes*. Vol. **3**. Springer.

Kumar, P. 2010. Probability distributions and the estimation of Ali-Mikhail-Haq copula. *Applied Mathematical Statistics*, **4**(14), 657–666.

Künsch, H. R. 2005. Recursive Monte Carlo filters: algorithms and theoretical analysis. *Annals of Statistics*, **33**(5), 1983–2021.

Kuon, S., Reich, A., & Reimers, L. 1987. The Cologne Re, Cologne. *ASTIN Bulletin*, **17**(2), 183.

Kuske, R., & Keller, J. B. 2000. Rate of convergence to a stable law. *SIAM Journal on Applied Mathematics*, **61**(4), 1308–1323.

Kusuoka, S. 2001. *On Law Invariant Coherent Risk Measures: Advances in Mathematical Economics*. Springer, 83–95.

Lagrange, J. L. 1770. *Nouvelle méthode pour résoudre les équations littérales par le moyen des séries*. Chez Haude et Spener, Libraires de la Cour & de l'Académie royale.

Landsman, Z. M., & Valdez, E. A. 2003. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, **7**(4), 55–71.

Landsman, Z., & Valdez, E. A. 2005. Tail conditional expectations for exponential dispersion models. *ASTIN Bulletin*, **35**(1), 189–209.

Landwehr, J. M., Matalas, N. C., & Wallis, J. R. 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, **15**(5), 1055–1064.

Laurie, D. 1997. Calculation of Gauss-Kronrod quadrature rules. *Mathematics of Computation of the American Mathematical Society*, **66**(219), 1133–1145.

Lavoie, J. L., Osler, T. J., & Tremblay, R. 1976. Fractional derivatives and special functions. *SIAM Review*, **18**(2), 240–268.

Lee, A. J. 1990. *U-Statistics: Theory and Practice*. Vol. **110**. CRC Press, 1–320.

Lee, M.-L. T., & Whitmore, G. A. 1993. Stochastic processes directed by randomized time. *Journal of Applied Probability*, **30**, 302–314.

Lehmann, E. L. 1966. Some concepts of dependence. *Annals of Mathematical Statistics*, **37**(1), 1137–1153.

Letac, G. 1989. Le problem de la classification des familles exponentielles naturelles de $\mathbb{R}^d$ ayant une fonction variance quadratique. In: *Probability Measures on Groups IX*. Springer, 192–216.

Lévy, P. 1924. Theorie des erreurs. La loi de Gauss et les Lois Exceptionelles. *Bulletin-Societe Mathematique de France*, **52**, 49–85.

Lévy, P. 1925. *Calcul des Probabilités*. Vol. **9**. Gauthier-Villars Paris.

Lévy, P. 1939. L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société Mathématique de France*, **67**, 1–41.

Lévy, P., & Borel, M. É. 1954. *Théorie de l'addition des variables aléatoires*. Vol. **1**. Gauthier-Villars Paris.

Li, M. F., Ren, J. R., & Zhu, T. 2009. *Series expansion in fractional calculus and fractional differential equations. Arxiv preprint arXiv:0910.4819*.

Li, J., & Tang, Q. 2010. A note on max-sum equivalence. *Statistics & Probability Letters*, **80**(23), 1720–1723.

Liu, J. S. 2008. *Monte Carlo Strategies in Scientific Computing*. Springer.

Liu, L. 2009. Precise large deviations for dependent random variables with heavy tails. *Statistics & Probability Letters*, **79**(9), 1290–1298.

Liu, J. S., & Chen, R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association*, **93**(443), 1032–1044.

Liu, J. S., Chen, R., & Wong, W. H. 1998. Rejection control and sequential importance sampling. *Journal of the American Statistical Association*, **93**(443), 1022–1031.

Lloyd, S. 1982. Least squares quantization in PCM. *Information Theory, IEEE Transactions on*, **28**(2), 129–137.

Lomax, K. S. 1954. Business failures: Another example of the analysis of failure data. *Journal of the American Statistical Association*, **49**(268), 847–852.

Lopes, H. F., Nascimento, F. F., & Gamerman, D. 2011. Generalized Pareto Models with Time-Varying Tail Behavior. Tech. Rept. Technical Report LES: UFRJ, in preparation.

Lu, J., Guo L. & Liu X. 2013. Measuring the operational risk of Chinese commercial banks using the semilinear credibility model. *Journal of Operational Risk*, **8**(2), 3–34.

Lugannani, R., & Rice, S. 1980. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, **12**(2), 475–490.

Lukacs, E. 1970. *Characteristic Functions*. Vol. **4**. London: Griffin.

Luo, X., & Shevchenko, P. V. 2011. A short tale of long tail integration. *Numerical Algorithms*, **56**(4), 577–590.

Macleod, A. J. 1989. A remark on algorithm AS 215: maximum-likelihood estimation of the parameters of the generalized extreme-value distribution. *Applied Statistics*, **38**(1), 198–199.

Madan, D. B., Carr, P. P., & Chang, E. C. 1998. The variance gamma process and option pricing. *European Finance Review*, **2**(1), 79–105.

Maejima, M., & Naito, Y. 1998. Semi-selfdecomposable distributions and a new class of limit theorems. *Probability Theory and Related Fields*, **112**(1), 13–31.

Mahbubul, M., Majumder, A., & Ali, M. M. 2008. A Comparison of methods of estimation of parameters of Tukey's gh family of distributions. *Pakistan Journal of Statistics*, **24**(2), 135–144.

Mainik, G., & Rüschendorf, L. 2010. On optimal portfolio diversification with respect to extreme risks. *Finance and Stochastics*, **14**(4), 593–623.

Mallor, F., & Omey, E. 2006. *Univariate and Multivariate Weighted Renewal Theory*. Vol. **2**. Collection of Monographies of the Department of Statistics and Operations Research.

Mandelbrot, B. 1960. The Pareto-Levey law and the distribution of income. *International Economic Review*, **1**(2), 79–106.

Mantegna, R. N., & Stanley, H. E. 1994. Stochastic process with ultraslow convergence to a Gaussian: the truncated Lévy flight. *Physical Review Letters*, **73**(22), 2946.

Marcus, M. B., & Rosiński, J. 2005. Continuity and boundedness of infinitely divisible processes: a Poisson point process approach. *Journal of Theoretical Probability*, **18**(1), 109–160.

Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. 2012. Approximate Bayesian computational methods. *Statistics and Computing*, **22**(6), 1167–1180.

Marinelli, C., D'addona, S., & Rachev, S. T. 2007. A comparison of some univariate models for Value-at-Risk and expected shortfall. *International Journal of Theoretical and Applied Finance*, **10**(06), 1043–1075.

Marinelli, C., & Rachev, S. 2000. Computational issues in stable financial modeling. *Applied Mathematics Reviews*, **1**, 285–327.

Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. 2003. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(26), 15324–15328.

Marshall, A. W., & Olkin, I. 1988. Families of multivariate distributions. *Journal of the American Statistical Association*, **83**(403), 834–841.

Martins, E. S., & Stedinger, J. R. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, **36**(3), 737–744.

Matacz, A. 2000. Financial modeling and option theory with the truncated Lévy process. *International Journal of Theoretical and Applied Finance*, **3**(01), 143–160.

Matthys, G., & Beirlant, J. 2003. Estimating the extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, **13**(3), 853–880.

Matuszewska, W. 1964. On a generalization of regularly increasing functions. *Studia Mathematica*, **24**(3), 271–279.

McCulloch, J. H. 1986. Simple consistent estimators of stable distribution parameters. *Communications in Statistics—Simulation and computation*, **15**(4), 1109–1136.

McCulloch, J. H. 1997. Measuring tail thickness to estimate the stable index α: a critique. *Journal of Business & Economic Statistics*, **15**(1), 74–81.

McCulloch, J. H. 1998. Linear regression with stable disturbances. In: *A Practical Guide to Heavy Tails, Statistical Techniques and Applications* (eds. R. Adler, R. Feldman, M. Taqqu). Birkhauser Boston Inc.: Cambridge, MA, 359–376.

McNeil, A. J. 1999. *Extreme Value Theory for Risk Managers*. Departement Mathematik ETH Zentrum.

McNeil, A. J., & Frey, R. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, **7**(3), 271–300.

McNeil, A. J., Frey, R., & Embrechts, P. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, NJ: Princeton University Press.

McNeil, A. J., & Nešlehová, J. 2009. Multivariate Archimedean Copulas, d-Monotone Functions and L_1-Norm Symmetric Distributions. *Annals of Statistics*, **37**(5), 3059–3097.

Melchiori, M. R. 2006. *Tools for sampling multivariate archimedian copulas*. www.YieldCurve.com.

Menn, C., & Rachev, S. T. 2006. Calibrated FFT-based density approximations for α-stable distributions. *Computational Statistics & Data Analysis*, **50**(8), 1891–1904.

Metzler, R., & Klafter, J. 2000. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Physics Reports*, **339**(1), 1–77.

Mignola, G., & Ugoccioni, R. 2006. Sources of uncertainty in modeling operational risk losses. *Journal of Operational Risk*, **1**(2), 33–50.

Mijnheer, J. L., & Mijnheer, J. L. 1975. *Sample Path Properties of Stable Processes*. Vol. **59**. Mathematisch Centrum Amsterdam.

Mikosch, T. 1999. *Regular Variation, Subexponentiality and their Applications in Probability Theory*. Eindhoven University of Technology.

Mikosch, T., & Nagaev, A. 2001. Rates in approximations to ruin probabilities for heavy-tailed distributions. *Extremes*, **4**(1), 67–78.

Mikosch, T., & Samorodnitsky, G. 2000. The supremum of a negative drift random walk with dependent heavy-tailed steps. *Annals of Applied Probability*, **10**(3), 1025–1064.

Miller, K. S., & Ross, B. 1993. *An Introduction to the Fractional Calculus and Fractional Differential Equations*. John Wiley & Sons.

von Mises, R. 1936. La distribution de la plus grande de n valeurs. *Revue de l'Union Interbalcanique*, **1**(1), 141–160.

Mittnik, S., Doganoglu, T., & Chenyao, D. 1999a. Computing the probability density function of the stable Paretian distribution. *Mathematical and Computer Modelling*, **29**(10), 235–240.

Mittnik, S., Rachev, S.T., Doganoglu, T., & Chenyao, D. 1999b. Maximum likelihood estimation of stable Paretian models. *Mathematical and Computer Modelling*, **29**(10), 275–293.

Mittnik, S., & Rachev, S. T. 1991. Alernative multivariate stable distributions and their applications to financial modelling. In: Cambanis, E., Samorodnitsky, G., & Taqqu, M. S. (eds), *Stable Processes and Related Topics*. Boston, MA: Birkhauser, 107–119.

Mohan, N. R., & Ravi, S. 1993. Max domains of attraction of univariate and multivariate p-max stable laws. *Theory of Probability & its Applications*, **37**(4), 632–643.

Mora, M. 1990. La convergence des fonctions variance des familles exponentielles naturelles. In: *Annales de la faculté des sciences de Toulouse*, Vol. **11**(2). Université Paul Sabatier, 105–120.

Morgenthaler, S., & Tukey, J. W. 2000. Fitting quantiles: Doubling, HR, HQ, and HHH distributions. *Journal of Computational and Graphical Statistics*, **9**(1), 180–195.

Morrison, J. E., & Smith, J. A. 2002. Stochastic modeling of flood peaks using the generalized extreme value distribution. *Water Resources Research*, **38**(12), 1305.

Moscadelli, M. 2004. *The modelling of operational risk: experiences with the analysis of the data collected by the Basel Committee*. Bank of Italy. working paper No. 517.

Mudholkar, G. S., & Hutson, A. D. 1998. LQ-moments: analogs of L-moments. *Journal of Statistical Planning and Inference*, **71**(1-2), 191–208.

Munir, A., & Shaw, W. 2012. *Quantile Mechanics 3: Series Representations and Approximation of some Quantile Functions appearing in Finance. arXiv preprint arXiv:1203.5729.*

Munkhammar, J. D. 2005. Fractional calculus and the Taylor–Riemann series. *Undergraduate Mathematics Journal*, **6**(1), 1–19.

Nam, D., & Gup, B. E. 2003. A quantile-fitting approach to value at risk for options. *Journal of Risk Finance*, **5**(1), 40–50.

Neal, R. M. 2001. Annealed importance sampling. *Statistics and Computing*, **11**(2), 125–139.

Necir, A., Rassoul, A., & Zitikis, R. 2010. Estimating the conditional tail expectation in the case of heavy-tailed losses. *Journal of Probability and Statistics*, **2010**, Article ID 596839, 17 pages, 2010. doi:10.1155/2010/596839

Nelsen, R. B. 1999. *An Introduction to Copulas*. Springer.

Nešlehová, J., Embrechts, P., & Chavez-Demoulin, V. 2006. Infinite mean models and the LDA for operational risk. *Journal of Operational Risk*, **1**(1), 3–25.

Newey, W. K., & Powell, J. L. 1987. Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, **55**(4), 819–847.

Ng, K. W., Tang, Q. H., & Yang, H. 2002. Maxima of sums of heavy-tailed random variables. *ASTIN Bulletin*, **32**(1), 43–56.

Nikias, C., & Shao, M. 1995. *Signal Processing with Alpha Stable Distributions and Applications*. Wiley, New York.

Nolan, J. P. 1997. Numerical calculation of stable densities and distribution functions. *Communications in Statistics—Stochastic Models*, **13**(4), 759–774.

Nolan, J. P. 1998. Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, **38**(2), 187–195.

Nolan, J. P. 1999. An algorithm for evaluating stable densities in Zolotarev's (iM) parameterization. *Mathematical and Computer Modelling*, **29**(10), 229–233.

Nolan, J. P. 2001. *Maximum likelihood estimation and diagnostics for stable distributions*. Lévy Processes. Springer, 379–400.

Nolan, J. 2015. *Stable Distributions: Models for Heavy-Tailed Data*. Birkhauser, Boston.

Novikov, E. A. 1994. Infinitely divisible distributions in turbulence. *Physical Review. E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics;(United States)*, **50**(5), R3303–R3305.

Oh, M. S., & Berger, J. O. 1993. Integration of multimodal functions by Monte Carlo importance sampling. *Journal of the American Statistical Association*, **88**(422), 450–456.

O'Hagan, A. 1998. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **47**(1), 21–35.

Oldham, K. B., & Spanier, J. 1974. *The Fractional Calculus: Theory and Applications of Differentiation and Integration to Arbitrary Order*. Vol. **111**. Elsevier Science.

Olshen, R. A., & Savage, L. J. 1969. *A Generalized Unimodality*. Defense Technical Information Center.

Omey, E., & Willekens, E. 1986. Second order behaviour of the tail of a subordinated probability distribution. *Stochastic Processes and their Applications*, **21**(2), 339–353.

Orey, S. 1968. On continuity properties of infinitely divisible distribution functions. *Annals of Mathematical Statistics*, **39**(3), 936–937.

Ortigueira, M. D. 2011. *Fractional Calculus for Scientists and Engineers*. Vol. **84**. Springer.

Ozaktas, H. M., Kutay, M. A., & Zalevsky, Z. 2001. *The Fractional Fourier Transform with Applications in Optics and Signal Processing*. New York: Wiley.

Palmer, K. J., Ridout, M. S., & Morgan, B. J. T. 2008. Modelling cell generation times by using the tempered stable distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **57**(4), 379–397.

Panjer, H. H. 1981. Recursive evaluation of a family of compound distribution. *ASTIN Bulletin*, **12**(1), 22–26.

Panjer, H. H. 2002. *Measureement of Risk, Solvency Requirements and Allocation of Capital Within Financial Conglomerates*. University of Waterloo, Institute of Insurance and Pension Research.

Panjer, H. H., & Wang, S. 1993. On the stability of recursive formulas. *ASTIN Bulletin*, **23**(2), 227–258.

Panjer, H. H., & Willmot, G. E. 1986. Computational aspects of recursive evaluation of compound distributions. *Insurance: Mathematics and Economics*, **5**(1), 113–116.

Panjer, H. H., Willmot, G. E., & Education, Actuarial. 1992. *Insurance Risk Models*. Society of Acturaries.

Pant, M. D., & Headrick, T. C. 2013. A method for simulating burr type III and type XII distributions through-moments and-correlations. *ISRN Applied Mathematics*, **2013**, Article ID 191604, 14 pages, doi:10.1155/2013/191604.

Pantcheva, E. 1985. Limit theorems for extreme order statistics under nonlinear normalization. *Stability Problems for Stochastic Models*. Springer, 284–309.

Pei, S.-C., & Hsue, W.-L. 2009. Random discrete fractional Fourier transform. *IEEE Signal Processing Letters*, **16**(12), 1015–1018.

Pei, S.-C., & Yeh, M.-H. 1997. Improved discrete fractional Fourier transform. *Optics Letters*, **22**(14), 1047–1049.

Peng, L., & Welsh, A. H. 2001. Robust estimation of the generalized pareto distribution. *Extremes*, **4**(1), 53–65.

Perez, N. H., Menendez, S. C., & Seco, L. 2003. *A theoretical comparison between moments and L-moments*. Unpublished paper retrieved from http://risklab. erin. utoronto. ca/members. htm.

Perreault, L., Bobée, B., & Rasmussen, P. F. 1999. Halphen distribution system. I: Mathematical and statistical properties. *Journal of Hydrologic Engineering*, **4**(3), 189–199.

Peters, G. W. 2005. *Topics in Sequential Monte Carlo Samplers*. MSc, University of Cambridge, Department of Engineering.

Peters, G. W., & Sisson, S. A. 2006. Bayesian inference, Monte Carlo sampling and operational risk. *Journal of Operational Risk*, **1**(3), 27–50.

Peters, G. W., Johansen, A. M., & Doucet, A. 2007. Simulation of the annual loss distribution in operational risk via Panjer recursions and Volterra integral equations for value-at-risk and expected shortfall estimation. *Journal of Operational Risk*, **2**(3), 29–58.

Peters, G. W., Shevchenko, P. V., & Wüthrich, M. V. 2009a. Dynamic operational risk: modeling dependence and combining different data sources of information. *Journal of Operational Risk*, **4**(2), 69–104.

Peters, G. W., Shevchenko, P. V., & Wüthrich, M. V. 2009b. Model uncertainty in claims reserving within Tweedie's compound Poisson models. *ASTIN Bulletin*, **39**(1), 1–33.

Peters, G. W., Fan, Y., & Sisson, S. A. 2012. On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. *Statistics and Computing*, **22**(6), 1209–1222.

Peters, G. W., Sisson, S. A., & Fan, Y. 2010a. Likelihood-free Bayesian inference for α-stable models. *Computational Statistics & Data Analysis*, **56**(1), 3743–3756.

Peters, G. W., Wüthrich, M. V., & Shevchenko, P. V. 2010b. Chain ladder method: Bayesian bootstrap versus classical bootstrap. *Insurance: Mathematics and Economics*, **47**(1), 36–51.

Peters, G. W., Shevchenko, P. V., Young, M., & Yip, W. 2011. Analytic loss distributional approach models for operational risk from the α-stable doubly stochastic compound processes and implications for capital allocation. *Insurance: Mathematics and Economics*, **49**(3), 565–579.

Peters, G. W., Fan, Y., & Sisson, S. A. 2012a. On sequential Monte Carlo, partial rejection control and approximate Bayesian computation. *Statistics and Computing*, **22**(6), 1209–1222.

Peters, G. W., Sisson, S. A., & Fan, Y. 2012b. Likelihood-free Bayesian inference for α-stable models. *Computational Statistics & Data Analysis*, **56**(11), 3743–3756.

Peters, G. W., Dong, A. X. D., & Kohn, R. 2014. A copula based Bayesian approach for paid incurred claims models for non-life insurance reserving. *Insurance: Mathematics and Economics*, **59**(0), 258–278.

Peters, G. W., Nevat, I., Septier, F., & Clavier, L. 2012d. *Generalized Interference Models in Doubly Stochastic Poisson Random Fields for Wideband Communications: the PNSC (alpha) model. arXiv preprint arXiv:1207.1531*.

Peters, G. W., Targino, R. S., & Shevchenko, P. V. 2013. Understanding operational risk capital approximations: first and second orders. *Governance and Regulation*, **2**(3), 1–30.

Petrov, V. V. 1975. *Sums of Independent Random Variables*. Berlin.

Pflug, G. Ch. 2000. Some remarks on the value-at-risk and the conditional value-at-risk. *Probabilistic Constrained Optimization*. Springer, 272–281.

Pitman, E. J. G. 1980. Subexponential distribution functions. *Journal of the Australian Mathematical Society Series A*, **29**(3), 337–347.

Podlubny, I. 1999. *Fractional Differential Equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to Methods of their Solution and Some of their Applications*. San Diego, CA: Academic Press.

Poirot, J., & Tankov, P. 2006. Monte Carlo option pricing for tempered stable (CGMY) processes. *Asia-Pacific Financial Markets*, **13**(4), 327–344.

Pollaczek, F. 1952. Fonctions caractéristiques de certaines répartitions définie au moyen de la notion d'ordre. *Application à la théorie des attentes (CR Acad. Se)*, **234**, a334.

Pollaczek, F. 1957. *Problèmes stochastiques posés par le phénomène de formation d'une queue d'attente à un guichet et par des phénomènes apparentés, Memorial des Sciences Mathematiques, fac. 136*.

Pollaczek, F. 1975. Order statistics of partial sums of mutually independent random variables. *Journal of Applied Probability*, **12**(2), 390–395.

Pollard, H. 1946. The representation of $\exp(-x^{\lambda})$ as a Laplace integral. *Bulletin of the American Mathematical Society*, **52**(10), 908–910.

Prescott, P., & Walden, A. T. 1980. Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika*, **67**(3), 723–724.

Press, S. J. 1972. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, **67**(340), 842–846.

Rachev, S., & Mittnik, S. 2000. *Stable Paretian Models in Finance*. New York: John Willey & Sons.

Rajput, B. S., & Rosiński, J. 1989. Spectral representations of infinitely divisible processes. *Probability Theory and Related Fields*, **82**(3), 451–487.

Raoult, J.-P., & Worms, R. 2003. Rate of convergence for the generalized Pareto approximation of the excesses. *Advances in Applied Probability*, **35**(4), 1007–1027.

Rayner, G. D., & MacGillivray, H. L. 2002. Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics and Computing*, **12**(1), 57–75.

Reeves, R. W., & Pettitt, A. N. 2005. A theoretical framework for approximate Bayesian computation. *Statistical Solutions to Modern Problems: Proceedings of the 20th International Workshop on Statistical Modelling*, Sydney, 393–396.

Reiss, R. D., & Thomas, M. 2007. *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and other Fields*. Birkhauser.

Resnick, S. I. 1986. Point processes, regular variation and weak convergence. *Advances in Applied Probability*, **18**(1), 66–138.

Resnick, S. I. 1997. Heavy tail modeling and teletraffic data: special invited paper. *Annals of Statistics*, **25**(5), 1805–1869.

Resnick, S. I. 2006. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Vol. **10**. Springer-Verlag.

Resnick, S. I. 2007. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag.

Richards, K.-A., Peters, G. W., & Dunsmuir, W. 2012. *Heavy-Tailed Features and Empirical Analysis of the Limit Order Book Volume Profiles in Futures Markets. arXiv preprint arXiv:1210.7215*.

Richardson, L. F. 1911. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, **210**, 307–357.

Ridout, M. S. 2009. Generating random numbers from a distribution specified by its Laplace transform. *Statistics and Computing*, **19**(4), 439–450.

Robert, C. Y., & Segers, J. 2008. Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance: Mathematics and Economics*, **43**(1), 85–92.

Roberts, G. E., & Kaufman, H. 1966. *Table of Laplace Transforms*. Saunders Philadelphia.

Rodriguez, R.N. 1977. A guide to the Burr type XII distributions. *Biometrika*, **64**(1), 129–134.

Rolin, J.-M., Jaumain, C., Embrechts, P., Walhin, J.-F., & Panjer, H. 2000. *Recursions for Actuaries and Applications in the Field of Reinsurance and Bonus-Malus Systems*. Universite Catholique de Louvain, Institut de Statistique, PhD thesis.

Rolski, T., Schmidli, H., Schmidt, V., & Teugels, J. 2009. *Stochastic Processes for Insurance and Finance*. Vol. **505**. Wiley. com.

Romberg, W. 1955. Vereinfachte numerische integration. *Det Kongelige Norske Videnskabers Selskab Forhandlinger*, **28**(7), 30–36.

Rootzén, H., & Tajvidi, N. 1997. Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal*, **1997**(1), 70–94.

Rosiński, J. 2001. Series representations of Lévy processes from the perspective of point processes. *Lévy Processes*. Springer, 401–415.

Rosiński, J. 2007. Tempering stable processes. *Stochastic Processes and their Applications*, **117**(6), 677–707.

Rosiński, J., & Samorodnitsky, G. 1993. Distributions of subadditive functionals of sample paths of infinitely divisible processes. *Annals of Probability*, **21**(2), 996–1014.

Rydberg, T. H. 1997. The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics. Stochastic Models*, **13**(4), 887–910.

Sahay, A., Wan, Z., & Keller, B. 2007. Operational risk capital: asymptotics in the case of heavy-tailed severity. *Journal of Operational Risk*, **2**(2), 61–72.

Samorodnitsky, G., & Taqqu, M. S. 1994a. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall/CRC.

Samorodnitsky, G., & Taqqu, M. S. 1994b. *Stable non-Gaussian processes*. New York: Chapman & Hall.

Samorodnitsky, G., & Taqqu, M. S. 1997. Stable non-Gaussian random processes. *Econometric Theory*, **13**(1), 133–142.

Sato, K. 1999. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press.

Satybaldina, K. I. 1972. Absolute estimates of the rate of convergence to stable laws. *Teoriya Veroyatnostei i ee Primeneniya*, **17**(4), 773–775.

Savu, C., & Trede, M. 2006. Hierarchical archimedean copulas. In: *International Conference on High Frequency Finance*, Konstanz, Germany.

Serfling, R. J. 1981. *Approximation Theorems of Mathematical Statistics (Wiley Series in Probability and Statistics)*. Wiley-Interscience.

Sgibnev, M. S. 1996. On the distribution of the maxima of partial sums. *Statistics & Probability Letters*, **28**(3), 235–238.

Shaw, W., Luu, T., & Brickman, N. 2011. *Quantile Mechanics II: Changes of Variables in Monte Carlo methods and GPU-Optimized Normal Quantiles. Available at SSRN 1969331.*

Shephard, N. G. 1991. From characteristic function to distribution function: a simple framework for the theory. *Econometric Theory*, **7**(4), 519–529.

Shevchenko, P. V. 2010. Implementing loss distribution approach for operational risk. *Applied Stochastic Models in Business and Industry*, **26**(3), 277–307.

Shevchenko, P. V. 2011. *Modelling Operational Risk Using Bayesian Inference*. Springer.

Shin Kim, Y., Rachev, S. T., Leonardo Bianchi, M., & Fabozzi, F. J. 2010. Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking & Finance*, **34**(9), 2096–2109.

Sichel, H. S. 1982. Repeat-buying and the generalized inverse Gaussian-Poisson distribution. *Applied Statistics*, **31**(3), 193–204.

Sillitto, G. P. 1969. Derivation of approximants to the inverse distribution function of a continuous univariate population from the order statistics of a sample. *Biometrika*, **56**(3), 641–650.

Singh, V. P., & Guo, H. 1995. Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME). *Hydrological Sciences Journal*, **40**(2), 165–181.

Sisson, S. A., Fan, Y., & Tanaka, M. M. 2007. Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(6), 1760–1765.

Sisson, S. A., Peters, G. W., Briers, M., & Fan, Y. 2010. *A note on target distribution ambiguity of likelihood-free samplers. arXiv preprint arXiv:1005.5201*.

Sisson, S. A., & Fan, Y. 2011. *Likelihood-Free MCMC.* Handbook of Markov Chain Monte Carlo, edited by Steve Brooks, Andrew Gelman, Galin Jones, Xiao-Li Meng, Chapman and Hall CRC Press, 313.

Skovgaard, I. B. M. 1987. Saddlepoint expansions for conditional distributions. *Journal of Applied Probability*, **24**(4), 875–887.

Slater, Lucy Joan 1960. *Confluent hypergeometric functions*, University Press Cambridge.

Small, C. G. 2010. *Expansions and Asymptotics for Statistics*. Vol. **115**. Chapman and Hall/CRC.

Smith, R. L. 1985. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**(1), 67–90.

Smith, R. L. 1986. Extreme value theory based on theır largest annual events. *Journal of Hydrology*, **86**(1), 27–43.

Smith, R. L. 1990. Extreme value theory. *Handbook of Applicable Mathematics*, W. Ledermann (ed.). John Wiley, Chichester, Chapter 14 ,437–472.

Soulier, P. 2009. *Some applications of regular variation in probability and statistics.* Escuela Venezolana de Matemáticas. Instituto Venezolano de Investigaciones Cientìficas, ISBN:978-980-261-111-9.

Stam, A. J. 1973. Regular variation of the tail of a subordinated probability distribution. *Advances in Applied Probability*, **5**(2), 308–327.

Steutel, F. W. 1973. Some recent results in infinite divisibility. *Stochastic Processes and their Applications*, **1**(2), 125–143.

Steutel, F. W., & Van Harn, K. 2003. *Infinite Divisibility of Probability Distributions on the Real Line*. CRC Press.

Sundt, B. 1998. A generalisation of the De Pril transform. *Scandinavian actuarial journal*, **1998**(1), 41–48.

Sundt, B. 2005. On some properties of De Pril transforms of counting distributions. *ASTIN Bulletin*, **25**(1), 19–31.

Sundt, B., & Jewell, W. S. 1981. Further results on recursive evaluation of compound distributions. *ASTIN Bulletin*, **12**(1), 27–39.

Sundt, B., & Vernic, R. 2009. *Recursions for Convolutions and Compound Distributions with Insurance Applications*. Springer: Berlin.

Tadikamalla, P. R. 1980. A look at the Burr and related distributions. *International Statistical Review/Revue Internationale de Statistique*, **48**(3), 337–344.

Takács, L. 1977. On the ordered partial sums of real random variables. *Journal of Applied Probability*, **14**(1), 75–88.

Tang, Q. 2006. Insensitivity to negative dependence of the asymptotic behavior of precise large deviations. *Electronic Journal of Probability*, **11**(4), 107–120.

Tang, Q. 2008. Insensitivity to negative dependence of asymptotic tail probabilities of sums and maxima of sums. *Stochastic Analysis and Applications*, **26**(3), 435–450.

Tarasov, V. E. 2008. Fractional vector calculus and fractional Maxwell?s equations. *Annals of Physics*, **323**(11), 2756–2778.

Tarov, V. A. 2004. Smoothly varying functions and perfect proximate orders. *Mathematical Notes*, **76**(1-2), 238–243.

Tawn, J. A. 1990. Modelling multivariate extreme value distributions. *Biometrika*, **77**(2), 245–253.

Taylor, J. W. 2008. Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, **6**(2), 231–252.

Teugels, J. L. 1975. The class of subexponential distributions. *Annals of Probability*, **3**(6), 1000–1011.

Teugels, J. L., & Veraverbeke, N. 1973. *Cramér-Type Estimates for the Probability of Ruin*. Research Institute Center for Statistics C.O.R.E. Discussion Paper No. 7316, available at: http://hdl.handle.net/1942/95.

Thorin, O. (1977) On the infinite divisibility of the lognormal distribution, *Scandinavian Actuarial Journal*, 1977 (3), 121–148.

Tong, B., & Wu, C. 2012. Asymptotics for operational risk quantified with a spectral risk measure. *Journal of Operational Risk*, **7**(3), 91.

Trinidad, W. I. 1990. A proof of Pollaczek-Spitzer identity. *International Journal of Mathematics and Mathematical Sciences*, **13**(4), 737–740.

Tukey, J. W. 1977a. *Exploratory Data Analysis*. Vol. **231**. Reading, MA; Addison-Wesley.

Tukey, J. W. 1977b. Modern techniques in data analysis. In: *NSF-Sponsored Regional Research Conference at Southern Massachusetts University (North Dartmouth, MA).*

Tweedie, M. C. K. 1947. Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proceedings of the Cambridge Philosophical Society*, Vol. **43**. Cambridge University Press, 100.

Tweedie, M. C. K. 1984. An index which distinguishes between some important exponential families. *Statistics: Applications and New Directions: Proc. Indian Statistical Institute Golden Jubilee International Conference*, 579–604.

Uchaikin, V. V., & Zolotarev, V. M. 1999. *Chance and Stability: Stable Distributions and their Applications*. Walter de Gruyter.

Usero, D. 2007. *Fractional Taylor Series for Caputo Fractional Derivatives. Construction of Numerical Schemes*. http://www. mat. ucm. es/deptos/ma/inv/prepub/new/2007-10. pdf.

Vinogradov, V. 1994. *Refined Large Deviation Limit Theorems*. Vol. **315**. CRC Press.

Waldmann, K.-H. 1996. Modified recursions for a class of compound distributions. *ASTIN Bulletin*, **26**(2), 213–224.

Walin, J. F., & Paris, J. 1998. On the use of equispaced discrete distributions. *ASTIN Bulletin*, **28**(2), 241–255.

Waller, L. A., Turnbull, B. W., & Hardin, J. M. 1995. Obtaining distribution functions by numerical inversion of characteristic functions with applications. *American Statistician*, **49**(4), 346–350.

Wang, Y., Cheng, D., & Wang, K. 2005. The closure of a local subexponential distribution class under convolution roots, with applications to the compound Poisson process. *Journal of Applied Probability*, **42**(4), 1194–1203.

Warde, W. D., & Katti, S. K. 1971. Infinite divisibility of discrete distributions, II. *Annals of Mathematical Statistics*, **42**(3), 1088–1090.

Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss?Newton method. *Biometrika*, **61**(3), 439–447.

Weissman, I. 1978. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, **73**(364), 812–815.

Wendel, J. G. 1960. Order statistics of partial sums. *The Annals of Mathematical Statistics*, **31**(4), 1034–1044.

Wendel, J. G. 1961. The non-absolute convergence of Gil-Pelaez'inversion integral. *Annals of Mathematical Statistics*, **32**(1), 338–339.

Weron, A. 1984. Stable processes and measures; a survey. In: *Probability Theory on Vector Spaces III*. Springer, 306–364.

Weron, R. 1996a. Correction To: "On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables". Tech. Rept. Hugo Steinhaus Center, Wroclaw University of Technology.

Weron, R. 1996b. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics & Probability Letters*, **28**(2), 165–171.

Weron, R. 2006. *Modeling and Forecasting Electiricty Loads and Prices: A Statistical Approach*. Wiley.

West, M. 1987. On scale mixtures of normal distributions. *Biometrika*, **74**(3), 646–648.

Willekens, E. 1989. Asymptotic approximations of compound distributions and some applications. *Bulletin de la Société Mathématique de Belgique. Série B*, **41**(1), 55–61.

Willekens, E., & Teugels, J. L. 1992. Asymptotic expansions for waiting time probabilities in an M/G/1 queue with long-tailed service time. *Queueing Systems*, **10**(4), 295–311.

Williams, D. 1991. *Probability with Martingales*. Cambridge University Press.

Wintner, A. 1936. On a class of Fourier transforms. *American Journal of Mathematics*, **58**(1), 45–90.

Wintner, A. 1938. *Lectures by Aurel Wintner on Asymptotic Distributions and Infinite Convolutions, 1937-1938*. Edwards Brothers, Inc.

Wintner, A. 1956. Cauchy's stable distributions and an "explicit formula" of Mellin. *American Journal of Mathematics*, **78**(4), 819–861.

Wright, E. M. 1935. The asymptotic expansion of the generalized Bessel function. *Proceedings of the London Mathematical Society*, **2**(1), 257–270.

Yamazato, M. 1978. Unimodality of infinitely divisible distribution functions of class L. *Annals of Probability*, **6**(4), 523–531.

Yu, J. 2004. Empirical characteristic function estimation and its applications. *Econometric Reviews*, **23**(2), 93–123.

Zolotarev, V. M. 1983. *Univariate Stable Distributions*. Moscow: Nauka.

Zolotarev, V. M. 1986. One-dimensional stable distributions. *Translations of Mathematical Monographs*. American Mathematical Society.

Zolotarev, V. M. 1994. On representation of densities of stable laws by special functions. *Theory of Probability & its Applications*, **39**(2), 354–362.

Index

α-Stable
  LDA models, 333
ν-Stable family, 332
  Tail properties, 332

Additive Exponential Dispersion Models, 312
Ali-Mikhail-Haq copula, 397
alpha-Stable distribution
  asymptotic density expansion, 196
  asymptotic tail behaviour, 220
  Characterization 1, 149
  Characterization 2, 149
  Characterization 3, 153
  Characterization 4, 156
  density approximation FFT and Bergstrom, 189
  density approximation polynomial series, 196
  density approximation quadrature, 187
  density approximation reparameterization, 198
  density Scale Mixture of Normals (SMiN), 206
  distribution approximation quadrature, 208
  distribution approximation series, 209
  domain of attraction, 152, 155
  Fractional Lower Order Moments, 223
  Levy distribution, 151
  Levy LePage series, 205
  Nolan's S0 and S1-Type Parameterization, 178
  parameter estimation, 215
  quantile function, 210
  simulation, 214
  Zolotarev's A-Type parameterization, 167, 172
  Zolotarev's B-Type density, 185
  Zolotarev's B-Type parameterization, 175
  Zolotarev's C and E-Type parameterizations, 176
  Zolotarev's W-Type parameterization, 176
alpha-Stable distribution spectral measure, 162
Archimedean copula, 397
Asymmetric Laplace distribution, 327
  Maximum-Likelihood estimation, 329
Asymmetric Power distribution, 446
Asymptotic equivalence in distribution, 110
asymptotically locally constant distribution, 120

Benktander II distribution, 68
Bergstrom series expansion, 196
Beta distribution, 592
Binomial-Levy LDA Model, 345
bounds
  aggregate tail distributions, 111
  probabilities of convolutions, 115
Box–Muller transformation, 66
Bromwich Inversion Integral for Laplace Transforms, 241
Burr distribution, 60
Burr Type III distribution, 61
Burr Type XII distribution, 60

Caputo fractional derivative, 513
Central Limit Theorem
  classical, 153
  Lindeberg condition, 154
  Lindeberg Levy, 153
  Lyapunov condition, 154

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.

© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

characteristic function, 140
comonotonicity and stochastic decreasing, 396
compound Poisson Levy process, 297
compound process
  tail decomposition, 415
consistent variation, 129
convolution, 282
convolution of distributions, 110
convolution root closure, 137
convolution semi-group, 281
  continous, 290
  generator, 290
  Levy process, 295
  truncated Poisson, 291
convolution symmetrization, 283
counting process, 75

De Bruyn conjugate, 126
De Pril's first method, 538
De Pril's second method, 539
Decoupage de Levy, 79
Discrete distributions
  D-distributions, 526
  extrapolation methods and acceleration, 533
  infinite divisibility, 527
  Linnik laws, 532
  non-degenerate, 525
  Panjer class, 525
  Sibuya distribution, 530
  Stable, 530
discretization of distribution, 519
distributions of common type, 25
domain of attraction, 145
  Stable and Tweedie convergence, 303
dominant variation, 130
doubly stochastic Binomial-Levy LDA model, 346
doubly stochastic Negative Binomial-Levy LDA model, 347
doubly stochastic Poisson-Levy LDA model, 345

Elongation transform, 255
equality in distribution, 21
Esscher transform, 229
expectiles, 478
Exponential Dispersion Family, 309, 441
  scale invariance, 313
Exponential Dispersion models
  infinite divisibility, 319
Exponential tilting, 229
extended and O-Type regular variation, 135, 359
extremal domain of attraction conditions, 33, 35
  Auxiliary function, 36
extremal limit distribution
  extremal limit distribution families, 39
Extreme Value Theory
  Block Maxima Approach, 40
  elemental percentile method, 95
  Generalized Extreme Value distribution, 26
  Generalized Pareto distribution (GPD), 56
  GEV distribution moments, 26, 54
  Gumbel family distributions, 65
  Maximum-likelihood GPD, 87
  mixed estimation methods MLE and L-Moments, 43
  Pareto distribution, 55
  Penultimate VaR approximation, 468
  Pickands estimator, 96
  POT's domain of attraction, 83
  Power-Normalized EVT domain of attraction, 469
  probability weighted moments GPD, 93
  profile likelihood GPD, 91
  simulation of GPD, 57
  small sample GPD, 90
  threshold exceedance, 72
Extreme value theory, 17
  Extremal Limit Problem, 25
Extreme Value Theory Peaks Over Threshold (POTs), 81

Faa Di Bruno's formula, 513
Farlie-Gumbel-Morgenstern copula, 401
Fast Fourier Transform, 189
First-Order stochastic dominance, 403
Fourier Inversion theorem, 181
fractional derivatives of composite functions, 513
Fractional Fourier Transform, 191
Fundamental Theorem of Calculus, 512

g-and-h distribution, 127, 257, 262
  ABC, 274
  index of regular variation, 268
  moments, 265
  percentile matching, 270
  sample L-moments, 271
  simulation, 262
  slow variation, 269

g-and-h distribution
L-moments, 261
Gaussian copula, 401
generalised inverse, 28, 452
properties, 28
generalized moments, 115
Gil–Pelaez inversion integral, 182
GPD small sample behaviour, 91

hazard function, 114
hazard rate, 114
higher order Panjer class, 573
Hinde–Demetrio frequency distributions, 325
Hougaard Levy domain of attraction, 307
Hougaard Levy process, 306
Hyper Geometric function, 203

Infinite Divisible, 144, 286, 293
absolute moments, 295
exponential moments, 295
fractional moments, 296
large deviations, 296
Lipschitz function, 296

Karamata's Representation Theorem, 125, 269, 487
Kolmogorov canonical characteristic function, 147
Kolmogorov three series theorem, 362

L-Class of distribution, 158
L-moment estimator, 273
L-moment Tukey transforms, 261
L-moments, 46
distribution based, 44
GEV method of moments, 49
sample estimators, 45
L-skewness and L-kurtosis
distribution, 47
Lagrange inversion formula, 212
Laguerre polynomials, 201
Landau notation
Big Oh and Little Oh, 106
large deviations inequality, 297
large sample asymptotics, 308
Levy canonical characteristic function, 147
Levy Cramer continuity theorem, 141
Levy measure, 294
Levy triplets, 161
Levy truncation functions, 162
local sub-exponential distribution, 120
locally heavy tailed, 128
log-concave density, 529
LogNormal distribution, 65
Long-Tailed distribution, 131
lower and upper negative association and dependence, 394
LQ-moments
distribution, 46
Lévy–Khintchine representation, 161

Matuszewska index, 137
max-sum equivalence, 112
mixed poisson distribution, 542
monotone density theorem, 481
multivariate negative dependence, 394

n-decomposable random variable, 143
n-divisible random variable, 144
n-fold convolution, 115
Natural Exponential Family, 305
mean and variance function, 305
Negative Binomial distribution, 103
Negative Binomial-Levy LDA model, 347
negative regression dependence, 395
Normal Exponential Family
steepness and regularity, 311
Normal Inverse Gaussian
semi-group, 288

O-regular variation distributions, 359
order statistics, 23

pairwise negative quadrant dependence, 395
pairwise positive quadrant dependence, 395
Panjer recursion, 567
Partial Rejection Control, 556
Pickands-Balkema-deHaan theorem, 73
Poisson summation of characteristic functions, 142
Poisson-Levy LDA model, 344
Poisson-Tweedie
discrete, 324
Poisson-Tweedie models, 309
Pollaczek–Spitzer–Wendell identity, 364
Polya's sufficient condition for characteristic functions, 141
Potter bounds, 126
probability distribution function, 21
probability generating function, 525
probability weighted moments, 94
profile likelihood, 53
deviance statistic, 53

quantile function, 29
GEV distribution, 31
properties, 29
tail, 29

recursions
continous Panjer recursion, 581
De Pril transform, 537
higher order recursions, 575
mixed Poisson, 577
Panjer recursion, 566
partial sums, 535
random sums, 565
Waldmann's recursion, 575
Wilmot class, 580
regular and slow variation, 121
quantile functions, 127
uniform convergence, 124
renewal process, 75
representation theorem of slowly varying functions, 431
reproductive Exponential Dispersion models, 313
return levels
Generalized Extreme Value distribution, 54
Richardson extrapolation, 533
risk measures
admissible risk spectrum, 492
CARA risk aversion function, 494
empirical quantile process, 450
ES for Asymmetric Power distribution, 447
ES for Exponential Dispersion Models, 443
ES for Exponential Dispersion Models Partial Sums, 444
ES for g-and-h, 446
EVT Hill estimator, 471
Extreme Value Theory penultimate approximation, 468
g-and-h model, 445
non-parametric estimators, 448
penultimate single-loss approximation POT's, 473
second order VaR, 458
single loss approximation VaR, 452
spectral risk measure single loss approximations, 478
TCE for Exponential Dispersion Models, 443
tempered, 503
VaR first and second-order single-loss approximations, 451
VaR for Asymmetric Power distribution, 447
VaR for Exponential Dispersion Models, 441
VaR for g-and-h, 445
VaR single-loss approximation with mean correction, 458
VaR time scaling property single loss approximation, 457
VaR via point processes, 473
Var, ES and spectral risk, 439
risk weighted assets, 438
Rosinski measure, 231

second order regular variation, 486
self-decomposable random variables, 145
Sequential Monte Carlo samplers, 550
severity distribution
splice model, 11
slowly varying distribution tail, 122
small sample asymptotics, 307
smoothly varying function, 129
splice models, 11
Stable rate of convergence
Stable rate of convergence, 225
Stieltjes integral, 157
sub-exponential distribution, 113, 118
subversively varying tail, 131

tail asymptotic
compound process light-tailed severity, 370
compound processes, 367
consistent variation, 373
convolution root closure, 358
convolution root closure O-regular varying, 360
dependent severity and frequency, 374
dominant frequency distribution, 372
first order single-loss approximation, 376
first order single-loss regular variation, 380
first-order single-loss sub-exponential, 377
Frechet domain of attraction, 373
generalized tail ratios, 381
Gumbel domain of attraction, 373
higher order single-loss approximation smooth variation, 432
higher order single-loss approximations, 414
higher order single-loss approximations slow variation, 431
independent frequency and severity, 375
Inverse Laplace Characters, 424
Laplace Characters, 423
large deviations for partial sums, 367

large number of losses, 362
long tailed distribution, 361
partial sum, 356
partial sum distributional lower bound, 357
partial sum long tailed, 362
partial sums o-regularly varying, 359
remainder analysis regular variation, 386
remainer analysis for sub-exponential, 385
Saddlepoint conditions, 370
second-order single-loss approximations, 389
second-order single-loss approximation dependent risks, 413
stochastic bounds for distributions of partial sums, 403
tail balance condition, 126
tail dependence measure, 398
tail distribution function, 21
tail equivalence, 21
tempered α-Stable LDA model, 231, 349
tempered Levy measure, 230
tempered Stable model
closure under convolution, 243
cummulants, 244, 245
density, 237, 241
Fractional Lower Order Moments, 243
Levy measure MTS, 239
parameter estimation, 246
representation 1, 232
representation 2 (CTS), 232
representation 3 (RDTS), 233
representation 4 (GTS), 234
representation 5 , 236
representation 6 (MTS), 239
representation 7 (NMTS), 240
representation CGMY, 235
shot noise representation, 249
simulation, 248
standardized, 245
tail behaviour, 252
Tier I Capital, 437
Tier I Capital Ratio, 438
Tier II Capital, 438
trimmed L-moments
distribution, 46
estimator variance, 48
sample estimators, 48
Tukey distribution, 268
Tukey transformation, 253
h-transform, 254
j-transform, 254
k-transform, 254
Tweedie convergence, 307
Tweedie loss model
cumulant function, 314
Tweedie models
parameter estimation, 317

underflow, 572
uniform convergence, 20
unimodality of distribution, 158

Vinogradov notation, 108
Volterra integral equations second kind, 581
von Mises condition
von Mises condition, 40

weak convergence in distribution
Helly–Bray theorem, 22
weak convergence in distribution, 21
weak convergence Levy process to Stable Levy process, 304
Weibull Stretched Exponential distribution, 119

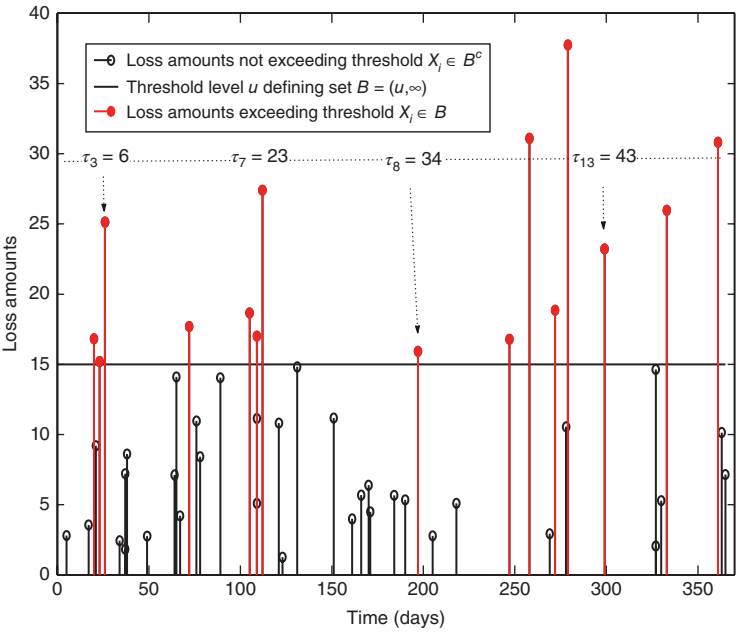

FIGURE 2.18 **An example realization of a single risk process $\{(X_i, t_i)\}_{i=1:n}$ with marked losses that exceeded a threshold loss amount u. The times on the x-axis at which the losses occurred correspond to the days t_i, the amounts correspond to the values X_i and the secondary process comprised of a subset of the losses, denoted by τ_j's as illustrated on the figure, corresponds to the indexes of the jth exceedance.**

Advances in Heavy Tailed Risk Modeling: A Handbook of Operational Risk,
First Edition. Gareth W. Peters and Pavel V. Shevchenko.
© 2015 John Wiley & Sons, Inc. Published 2015 by John Wiley & Sons, Inc.

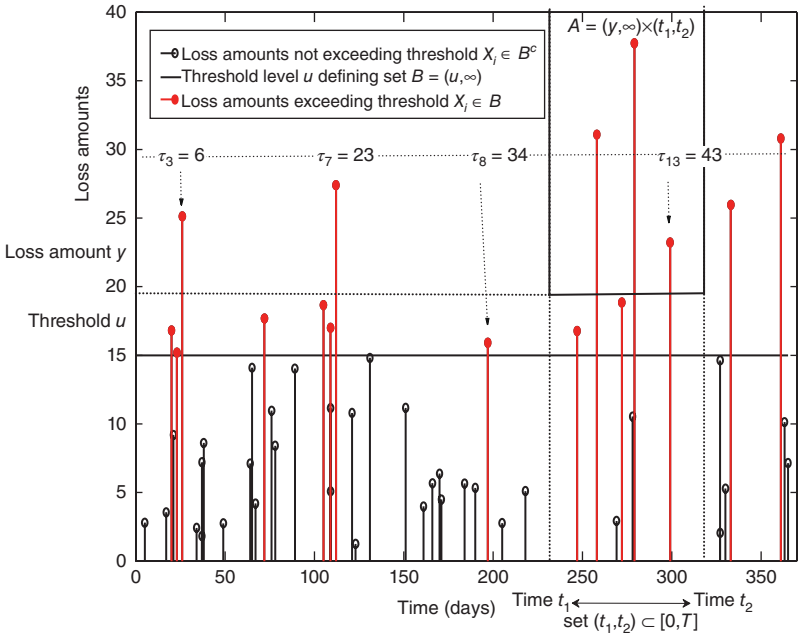

FIGURE 2.19 An example realization of a single risk process $\{(X_i, t_i)\}_{i=1:n}$ with marked losses that exceeded a threshold loss amount u. The times on the x-axis at which the losses occurred correspond to the days t_i, the amounts correspond to the values X_i and the secondary process comprised of a subset of the losses, denoted by τ_j's as illustrated on the figure, correspond to the indexes of the jth exceedance. In addition, an example region $A = (t_1, t_2) \times (y, \infty)$ is marked for points exceeding threshold u in time interval $(t_1, t_2) \subset [0, T]$.

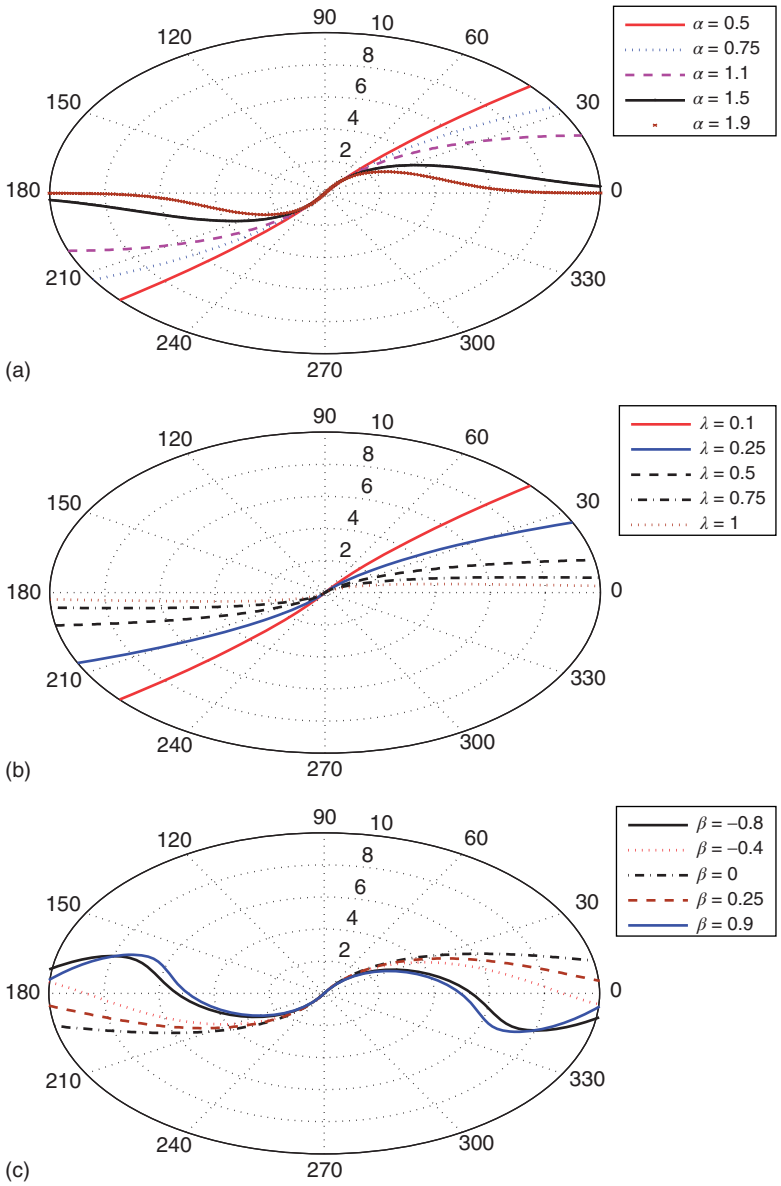

FIGURE 4.2 (a) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\alpha \in \{0.5, 0.75, 1.1, 1.5, 1.9\}$ with $\beta = 0$, $\gamma = 0.1$ and $\delta = 0$. (b) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\gamma \in \{0.1, 0.25, 0.5, 0.75, 1\}$ with $\alpha = 0.5$, $\beta = 0$ and $\delta = 0$. (c) Example of the A-type parameterization of the α-stable characteristic function for a range of values of $\beta \in \{-0.8, -0.4, 0, 0.25, 0.9\}$ with $\alpha = 1.2$, $\gamma = 0.1$ and $\delta = 0$.

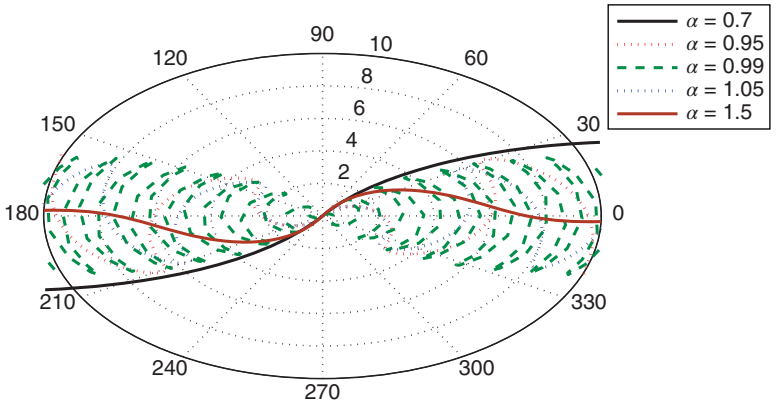

FIGURE 4.3 Example of A-type paramerization for a range of stability index values.

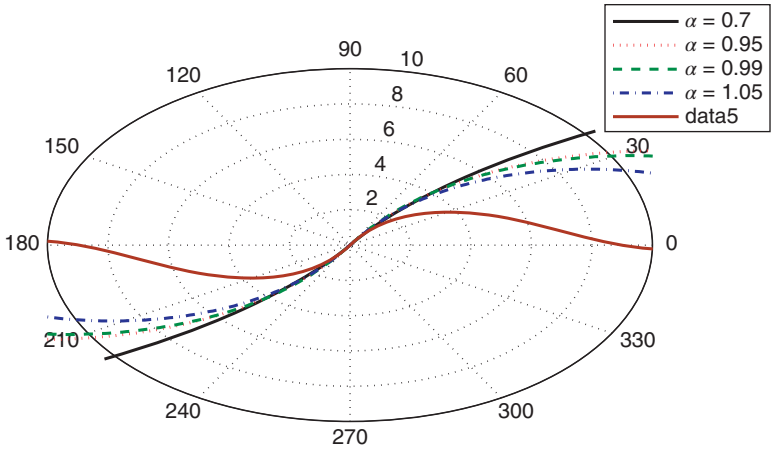

FIGURE 4.4 Example of B-type paramerization for a range of stability index values.

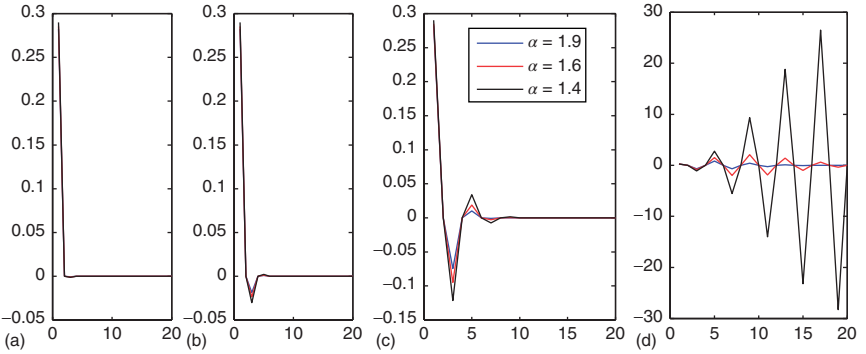

FIGURE 4.9 Study of first 20 summand terms of the stable density series expansion for a range of parameter values at three different x locations. (a) $x = .1$, (b) $x = 0.5$, (c) $x = 1$ and (d) $x = 3$.

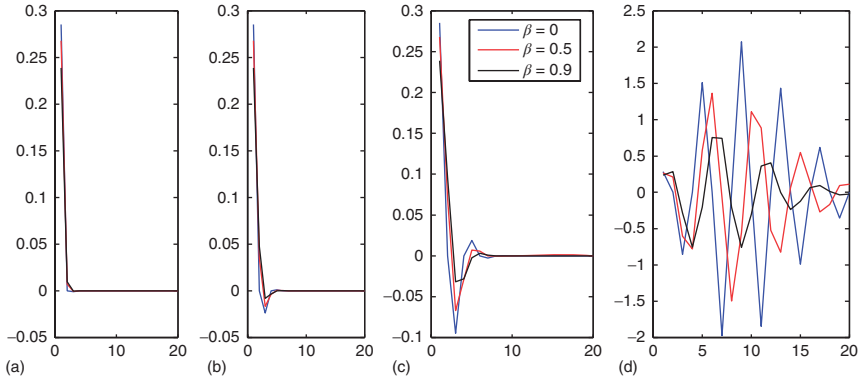

FIGURE 4.10 Study of first 20 summand terms of the stable density series expansion for a range of parameter values at three different x locations. (a) $x = .1$, (b) $x = 0.5$, (c) $x = 1$ and (d) $x = 3$.

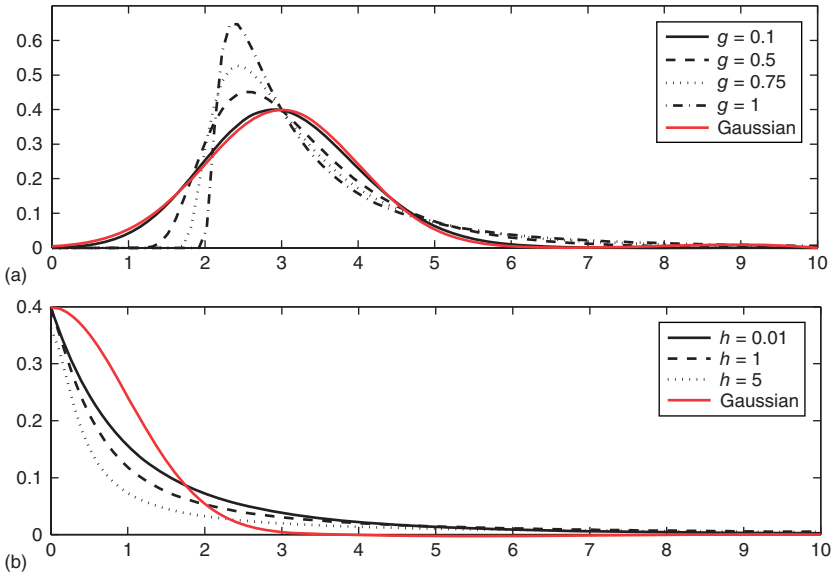

FIGURE 5.7 (a) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base Gaussian distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (b) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base Gaussian distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$.

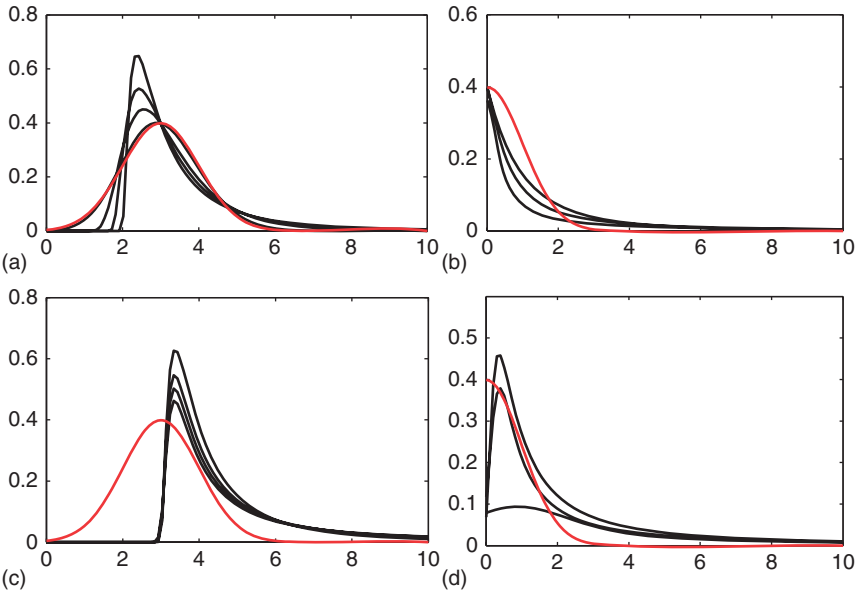

FIGURE 5.8 **(a) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base Gaussian distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (b) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base Gaussian distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$. (c) This plot shows the effect of the skewness parameter g on the elongation-transformed severity distribution versus the base LogNormal(0,1) distribution with $g \in \{0.1, 0.5, 0.75, 1\}$. In this case, the other parameters were set to $a = 3$, $b = 1$ and $h = 0.001$. (d) This plot shows the effect of the kurtosis parameter h on the elongation-transformed severity distribution versus the base LogNormal(0,1) distribution with $h \in \{0.01, 1, 5\}$. In this case, the other parameters were set to $a = 0$, $b = 1$ and $g = 1$.**

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley’s ebook EULA.