Reliability Engineering

Probabilistic Models and Maintenance Methods

MECHANICAL ENGINEERING
A Series of Textbooks and Reference Books

Founding Editor

L. L. Faulkner

Columbus Division, Battelle Memorial Institute
and Department of Mechanical Engineering
The Ohio State University
Columbus, Ohio

1. *Spring Designer's Handbook*, Harold Carlson
2. *Computer-Aided Graphics and Design*, Daniel L. Ryan
3. *Lubrication Fundamentals*, J. George Wills
4. *Solar Engineering for Domestic Buildings*, William A. Himmelman
5. *Applied Engineering Mechanics: Statics and Dynamics*, G. Boothroyd and C. Poli
6. *Centrifugal Pump Clinic*, Igor J. Karassik
7. *Computer-Aided Kinetics for Machine Design*, Daniel L. Ryan
8. *Plastics Products Design Handbook, Part A: Materials and Components; Part B: Processes and Design for Processes*, edited by Edward Miller
9. *Turbomachinery: Basic Theory and Applications*, Earl Logan, Jr.
10. *Vibrations of Shells and Plates*, Werner Soedel
11. *Flat and Corrugated Diaphragm Design Handbook*, Mario Di Giovanni
12. *Practical Stress Analysis in Engineering Design*, Alexander Blake
13. *An Introduction to the Design and Behavior of Bolted Joints*, John H. Bickford
14. *Optimal Engineering Design: Principles and Applications*, James N. Siddall
15. *Spring Manufacturing Handbook*, Harold Carlson
16. *Industrial Noise Control: Fundamentals and Applications*, edited by Lewis H. Bell
17. *Gears and Their Vibration: A Basic Approach to Understanding Gear Noise*, J. Derek Smith
18. *Chains for Power Transmission and Material Handling: Design and Applications Handbook*, American Chain Association
19. *Corrosion and Corrosion Protection Handbook*, edited by Philip A. Schweitzer
20. *Gear Drive Systems: Design and Application*, Peter Lynwander
21. *Controlling In-Plant Airborne Contaminants: Systems Design and Calculations*, John D. Constance
22. *CAD/CAM Systems Planning and Implementation*, Charles S. Knox
23. *Probabilistic Engineering Design: Principles and Applications*, James N. Siddall

24. *Traction Drives: Selection and Application*, Frederick W. Heilich III and Eugene E. Shube
25. *Finite Element Methods: An Introduction*, Ronald L. Huston and Chris E. Passerello
26. *Mechanical Fastening of Plastics: An Engineering Handbook*, Brayton Lincoln, Kenneth J. Gomes, and James F. Braden
27. *Lubrication in Practice: Second Edition*, edited by W. S. Robertson
28. *Principles of Automated Drafting*, Daniel L. Ryan
29. *Practical Seal Design*, edited by Leonard J. Martini
30. *Engineering Documentation for CAD/CAM Applications*, Charles S. Knox
31. *Design Dimensioning with Computer Graphics Applications*, Jerome C. Lange
32. *Mechanism Analysis: Simplified Graphical and Analytical Techniques*, Lyndon O. Barton
33. *CAD/CAM Systems: Justification, Implementation, Productivity Measurement*, Edward J. Preston, George W. Crawford, and Mark E. Coticchia
34. *Steam Plant Calculations Manual*, V. Ganapathy
35. *Design Assurance for Engineers and Managers*, John A. Burgess
36. *Heat Transfer Fluids and Systems for Process and Energy Applications*, Jasbir Singh
37. *Potential Flows: Computer Graphic Solutions*, Robert H. Kirchhoff
38. *Computer-Aided Graphics and Design: Second Edition*, Daniel L. Ryan
39. *Electronically Controlled Proportional Valves: Selection and Application*, Michael J. Tonyan, edited by Tobi Goldoftas
40. *Pressure Gauge Handbook*, AMETEK, U.S. Gauge Division, edited by Philip W. Harland
41. *Fabric Filtration for Combustion Sources: Fundamentals and Basic Technology*, R. P. Donovan
42. *Design of Mechanical Joints*, Alexander Blake
43. *CAD/CAM Dictionary*, Edward J. Preston, George W. Crawford, and Mark E. Coticchia
44. *Machinery Adhesives for Locking, Retaining, and Sealing*, Girard S. Haviland
45. *Couplings and Joints: Design, Selection, and Application*, Jon R. Mancuso
46. *Shaft Alignment Handbook*, John Piotrowski
47. *BASIC Programs for Steam Plant Engineers: Boilers, Combustion, Fluid Flow, and Heat Transfer*, V. Ganapathy
48. *Solving Mechanical Design Problems with Computer Graphics*, Jerome C. Lange
49. *Plastics Gearing: Selection and Application*, Clifford E. Adams
50. *Clutches and Brakes: Design and Selection*, William C. Orthwein
51. *Transducers in Mechanical and Electronic Design*, Harry L. Trietley
52. *Metallurgical Applications of Shock-Wave and High-Strain-Rate Phenomena*, edited by Lawrence E. Murr, Karl P. Staudhammer, and Marc A. Meyers
53. *Magnesium Products Design*, Robert S. Busk
54. *How to Integrate CAD/CAM Systems: Management and Technology*, William D. Engelke

55. *Cam Design and Manufacture: Second Edition; with cam design software for the IBM PC and compatibles*, disk included, Preben W. Jensen
56. *Solid-State AC Motor Controls: Selection and Application*, Sylvester Campbell
57. *Fundamentals of Robotics*, David D. Ardayfio
58. *Belt Selection and Application for Engineers*, edited by Wallace D. Erickson
59. *Developing Three-Dimensional CAD Software with the IBM PC*, C. Stan Wei
60. *Organizing Data for CIM Applications*, Charles S. Knox, with contributions by Thomas C. Boos, Ross S. Culverhouse, and Paul F. Muchnicki
61. *Computer-Aided Simulation in Railway Dynamics*, by Rao V. Dukkipati and Joseph R. Amyot
62. *Fiber-Reinforced Composites: Materials, Manufacturing, and Design*, P. K. Mallick
63. *Photoelectric Sensors and Controls: Selection and Application*, Scott M. Juds
64. *Finite Element Analysis with Personal Computers*, Edward R. Champion, Jr. and J. Michael Ensminger
65. *Ultrasonics: Fundamentals, Technology, Applications: Second Edition, Revised and Expanded*, Dale Ensminger
66. *Applied Finite Element Modeling: Practical Problem Solving for Engineers*, Jeffrey M. Steele
67. *Measurement and Instrumentation in Engineering: Principles and Basic Laboratory Experiments*, Francis S. Tse and Ivan E. Morse
68. *Centrifugal Pump Clinic: Second Edition, Revised and Expanded*, Igor J. Karassik
69. *Practical Stress Analysis in Engineering Design: Second Edition, Revised and Expanded*, Alexander Blake
70. *An Introduction to the Design and Behavior of Bolted Joints: Second Edition, Revised and Expanded*, John H. Bickford
71. *High Vacuum Technology: A Practical Guide*, Marsbed H. Hablanian
72. *Pressure Sensors: Selection and Application*, Duane Tandeske
73. *Zinc Handbook: Properties, Processing, and Use in Design*, Frank Porter
74. *Thermal Fatigue of Metals*, Andrzej Weronski and Tadeusz Hejwowski
75. *Classical and Modern Mechanisms for Engineers and Inventors*, Preben W. Jensen
76. *Handbook of Electronic Package Design*, edited by Michael Pecht
77. *Shock-Wave and High-Strain-Rate Phenomena in Materials*, edited by Marc A. Meyers, Lawrence E. Murr, and Karl P. Staudhammer
78. *Industrial Refrigeration: Principles, Design and Applications*, P. C. Koelet
79. *Applied Combustion*, Eugene L. Keating
80. *Engine Oils and Automotive Lubrication*, edited by Wilfried J. Bartz
81. *Mechanism Analysis: Simplified and Graphical Techniques, Second Edition, Revised and Expanded*, Lyndon O. Barton
82. *Fundamental Fluid Mechanics for the Practicing Engineer*, James W. Murdock
83. *Fiber-Reinforced Composites: Materials, Manufacturing, and Design, Second Edition, Revised and Expanded*, P. K. Mallick
84. *Numerical Methods for Engineering Applications*, Edward R. Champion, Jr.

85. *Turbomachinery: Basic Theory and Applications, Second Edition, Revised and Expanded*, Earl Logan, Jr.
86. *Vibrations of Shells and Plates: Second Edition, Revised and Expanded*, Werner Soedel
87. *Steam Plant Calculations Manual: Second Edition, Revised and Expanded*, V. Ganapathy
88. *Industrial Noise Control: Fundamentals and Applications, Second Edition, Revised and Expanded*, Lewis H. Bell and Douglas H. Bell
89. *Finite Elements: Their Design and Performance*, Richard H. MacNeal
90. *Mechanical Properties of Polymers and Composites: Second Edition, Revised and Expanded*, Lawrence E. Nielsen and Robert F. Landel
91. *Mechanical Wear Prediction and Prevention*, Raymond G. Bayer
92. *Mechanical Power Transmission Components*, edited by David W. South and Jon R. Mancuso
93. *Handbook of Turbomachinery*, edited by Earl Logan, Jr.
94. *Engineering Documentation Control Practices and Procedures*, Ray E. Monahan
95. *Refractory Linings Thermomechanical Design and Applications*, Charles A. Schacht
96. *Geometric Dimensioning and Tolerancing: Applications and Techniques for Use in Design, Manufacturing, and Inspection*, James D. Meadows
97. *An Introduction to the Design and Behavior of Bolted Joints: Third Edition, Revised and Expanded*, John H. Bickford
98. *Shaft Alignment Handbook: Second Edition, Revised and Expanded*, John Piotrowski
99. *Computer-Aided Design of Polymer-Matrix Composite Structures*, edited by Suong Van Hoa
100. *Friction Science and Technology*, Peter J. Blau
101. *Introduction to Plastics and Composites: Mechanical Properties and Engineering Applications*, Edward Miller
102. *Practical Fracture Mechanics in Design*, Alexander Blake
103. *Pump Characteristics and Applications*, Michael W. Volk
104. *Optical Principles and Technology for Engineers*, James E. Stewart
105. *Optimizing the Shape of Mechanical Elements and Structures*, A. A. Seireg and Jorge Rodriguez
106. *Kinematics and Dynamics of Machinery*, Vladimír Stejskal and Michael Valásek
107. *Shaft Seals for Dynamic Applications*, Les Horve
108. *Reliability-Based Mechanical Design*, edited by Thomas A. Cruse
109. *Mechanical Fastening, Joining, and Assembly*, James A. Speck
110. *Turbomachinery Fluid Dynamics and Heat Transfer*, edited by Chunill Hah
111. *High-Vacuum Technology: A Practical Guide, Second Edition, Revised and Expanded*, Marsbed H. Hablanian
112. *Geometric Dimensioning and Tolerancing: Workbook and Answerbook*, James D. Meadows
113. *Handbook of Materials Selection for Engineering Applications,* edited by G. T. Murray

114. *Handbook of Thermoplastic Piping System Design*, Thomas Sixsmith and Reinhard Hanselka
115. *Practical Guide to Finite Elements: A Solid Mechanics* Approach, Steven M. Lepi
116. *Applied Computational Fluid Dynamics*, edited by Vijay K. Garg
117. *Fluid Sealing Technology*, Heinz K. Muller and Bernard S. Nau
118. *Friction and Lubrication in Mechanical Design*, A. A. Seireg
119. *Influence Functions and Matrices*, Yuri A. Melnikov
120. *Mechanical Analysis of Electronic Packaging Systems*, Stephen A. McKeown
121. *Couplings and Joints: Design, Selection, and Application, Second Edition, Revised and Expanded*, Jon R. Mancuso
122. *Thermodynamics: Processes and Applications*, Earl Logan, Jr.
123. *Gear Noise and Vibration*, J. Derek Smith
124. *Practical Fluid Mechanics for Engineering Applications*, John J. Bloomer
125. *Handbook of Hydraulic Fluid Technology*, edited by George E. Totten
126. *Heat Exchanger Design Handbook*, T. Kuppan
127. *Designing for Product Sound Quality*, Richard H. Lyon
128. *Probability Applications in Mechanical Design*, Franklin E. Fisher and Joy R. Fisher
129. *Nickel Alloys*, edited by Ulrich Heubner
130. *Rotating Machinery Vibration: Problem Analysis and Troubleshooting*, Maurice L. Adams, Jr.
131. *Formulas for Dynamic Analysis*, Ronald L. Huston and C. Q. Liu
132. *Handbook of Machinery Dynamics*, Lynn L. Faulkner and Earl Logan, Jr.
133. *Rapid Prototyping Technology: Selection and Application*, Kenneth G. Cooper
134. *Reciprocating Machinery Dynamics: Design and Analysis*, Abdulla S. Rangwala
135. *Maintenance Excellence: Optimizing Equipment Life-Cycle Decisions*, edited by John D. Campbell and Andrew K. S. Jardine
136. *Practical Guide to Industrial Boiler Systems*, Ralph L. Vandagriff
137. *Lubrication Fundamentals: Second Edition, Revised and Expanded*, D. M. Pirro and A. A. Wessol
138. *Mechanical Life Cycle Handbook: Good Environmental Design and Manufacturing*, edited by Mahendra S. Hundal
139. *Micromachining of Engineering Materials*, edited by Joseph McGeough
140. *Control Strategies for Dynamic Systems: Design and Implementation*, John H. Lumkes, Jr.
141. *Practical Guide to Pressure Vessel Manufacturing*, Sunil Pullarcot
142. *Nondestructive Evaluation: Theory, Techniques, and Applications*, edited by Peter J. Shull
143. *Diesel Engine Engineering: Thermodynamics, Dynamics, Design, and Control*, Andrei Makartchouk
144. *Handbook of Machine Tool Analysis*, Ioan D. Marinescu, Constantin Ispas, and Dan Boboc

145. *Implementing Concurrent Engineering in Small Companies*, Susan Carlson Skalak
146. *Practical Guide to the Packaging of Electronics: Thermal and Mechanical Design and Analysis*, Ali Jamnia
147. *Bearing Design in Machinery: Engineering Tribology and Lubrication*, Avraham Harnoy
148. *Mechanical Reliability Improvement: Probability and Statistics for Experimental Testing*, R. E. Little
149. *Industrial Boilers and Heat Recovery Steam Generators: Design, Applications, and Calculations*, V. Ganapathy
150. *The CAD Guidebook: A Basic Manual for Understanding and Improving Computer-Aided Design*, Stephen J. Schoonmaker
151. *Industrial Noise Control and Acoustics*, Randall F. Barron
152. *Mechanical Properties of Engineered Materials*, Wolé Soboyejo
153. *Reliability Verification, Testing, and Analysis in Engineering Design*, Gary S. Wasserman
154. *Fundamental Mechanics of Fluids: Third Edition*, I. G. Currie
155. *Intermediate Heat Transfer*, Kau-Fui Vincent Wong
156. *HVAC Water Chillers and Cooling Towers: Fundamentals, Application, and Operation*, Herbert W. Stanford III
157. *Gear Noise and Vibration: Second Edition, Revised and Expanded*, J. Derek Smith
158. *Handbook of Turbomachinery: Second Edition*, *Revised and Expanded*, edited by Earl Logan, Jr. and Ramendra Roy
159. *Piping and Pipeline Engineering: Design, Construction, Maintenance, Integrity, and Repair*, George A. Antaki
160. *Turbomachinery: Design and Theory*, Rama S. R. Gorla and Aijaz Ahmed Khan
161. *Target Costing: Market-Driven Product Design*, M. Bradford Clifton, Henry M. B. Bird, Robert E. Albano, and Wesley P. Townsend
162. *Fluidized Bed Combustion*, Simeon N. Oka
163. *Theory of Dimensioning: An Introduction to Parameterizing Geometric Models*, Vijay Srinivasan
164. *Handbook of Mechanical Alloy Design*, edited by George E. Totten, Lin Xie, and Kiyoshi Funatani
165. *Structural Analysis of Polymeric Composite Materials*, Mark E. Tuttle
166. *Modeling and Simulation for Material Selection and Mechanical Design*, edited by George E. Totten, Lin Xie, and Kiyoshi Funatani
167. *Handbook of Pneumatic Conveying Engineering*, David Mills, Mark G. Jones, and Vijay K. Agarwal
168. *Clutches and Brakes: Design and Selection, Second Edition*, William C. Orthwein
169. *Fundamentals of Fluid Film Lubrication: Second Edition*, Bernard J. Hamrock, Steven R. Schmid, and Bo O. Jacobson
170. *Handbook of Lead-Free Solder Technology for Microelectronic Assemblies*, edited by Karl J. Puttlitz and Kathleen A. Stalter
171. *Vehicle Stability*, Dean Karnopp

172. *Mechanical Wear Fundamentals and Testing: Second Edition, Revised and Expanded*, Raymond G. Bayer
173. *Liquid Pipeline Hydraulics*, E. Shashi Menon
174. *Solid Fuels Combustion and Gasification*, Marcio L. de Souza-Santos
175. *Mechanical Tolerance Stackup and Analysis*, Bryan R. Fischer
176. *Engineering Design for Wear,* Raymond G. Bayer
177. *Vibrations of Shells and Plates: Third Edition, Revised and Expanded*, Werner Soedel
178. *Refractories Handbook*, edited by Charles A. Schacht
179. *Practical Engineering Failure Analysis*, Hani M. Tawancy, Anwar Ul-Hamid, and Nureddin M. Abbas
180. *Mechanical Alloying and Milling*, C. Suryanarayana
181. *Mechanical Vibration: Analysis, Uncertainties, and Control, Second Edition, Revised and Expanded*, Haym Benaroya
182. *Design of Automatic Machinery*, Stephen J. Derby
183. *Practical Fracture Mechanics in Design: Second Edition, Revised and Expanded*, Arun Shukla
184. *Practical Guide to Designed Experiments*, Paul D. Funkenbusch
185. *Gigacycle Fatigue in Mechanical Practive*, Claude Bathias and Paul C. Paris
186. *Selection of Engineering Materials and Adhesives*, Lawrence W. Fisher
187. *Boundary Methods: Elements, Contours, and Nodes*, Subrata Mukherjee and Yu Xie Mukherjee
188. *Rotordynamics*, Agnieszka Muszynska
189. *Pump Characteristics and Applications: Second Edition*, Michael Volk
190. *Reliability Engineering: Probabilistic Models and Maintenance Methods*, Joel A. Nachlas

Reliability Engineering

Probabilistic Models and Maintenance Methods

Joel A. Nachlas

Taylor & Francis
Taylor & Francis Group

Boca Raton London New York Singapore

A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20110713

International Standard Book Number-13: 978-1-4200-5618-1 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedicated to the memory of Betty Nachlas

Preface

The motivation for the preparation of this book is my wish to create an integrated introductory resource for the study of reliability evaluation and maintenance planning. The focus across all of the topics treated is the support of design activities that lead to the production of dependable and efficient equipment. The orientation of the topical development is that probability models provide an effective vehicle for portraying and evaluating the variability that is inherent in the performance and longevity of equipment.

The book is intended to support either an introductory course in reliability theory and preventive maintenance planning or a sequence of courses that address these topics. Fairly comprehensive coverage of the basic models and of various methods of analysis is provided. An understanding of the topics discussed should permit the reader to comprehend the literature describing new and advanced models and methods.

Notwithstanding the emphasis upon initial study, the text should also serve well as a resource for practicing engineers. Engineers who are involved in the design process should find a coherent explanation of the reliability and maintenance issues that will influence the success of the devices they create. Similarly, engineers responsible for the analysis and verification of product reliability or for the planning of maintenance support of fielded equipment should find the material presented here to be relevant and easy to access and use.

The background required of the reader is a sound understanding of probability. This subsumes capability with calculus. More specifically, the reader should have an understanding of distribution theory, Laplace transforms, convolutions, stochastic processes, and Markov processes. It is also worth mentioning that the use of the methods discussed in this book often involves substantial computational effort, so facility with numerical methods and access to efficient mathematical software is desirable.

One caveat concerning the coverage here is that the treatment is strictly limited to hardware. Reliability and maintenance models have been developed for applications to software, humans, and services systems. No criticism of those efforts is intended, but the focus here is simply hardware.

The organization of the text is reasonably straightforward. The elementary concepts of reliability theory are presented sequentially in Chapters 1 through 6. Following this, commonly used statistical methods for evaluating component reliability are described in Chapters 7 and 8. Chapters 9 through 13 treat repairable systems and thus maintenance planning models. Here again, the presentation is sequential in that simple failure models precede those that include preventive actions, and the renewal cases are treated before the more realistic nonrenewal cases. In the final chapter, four interesting special topics, including warranties, are discussed. It is worth noting that three appendices that address aspects of numerical computation are provided. These should be quite useful to the reader.

Naturally, many people have contributed to the preparation of this text. The principal factor in the completion of this book has been the support and encouragement of my wife Beverley. An important practical component of my success has been the support of Virginia Tech, particularly in providing computing resources and time during my sabbatical. I also wish to express my profound gratitude to my graduate students who have taught me so much about these topics over the years. May we all continue to learn and grow and to enjoy the study of this important subject.

About the Author

Joel A. Nachlas serves on the faculty of the Grado Department of Industrial and Systems Engineering at Virginia Tech. He has taught at Virginia Tech since 1974 and acts as the coordinator for the department's graduate program in Operations Research. For the past 12 years, he has also taught Reliability Theory regularly at the Ecole Superiore d'Ingenieures de Nice-Sophia Antipolis.

Dr. Nachlas earned his B.E.S. from the Johns Hopkins University in 1970. He earned his M.S. in 1972. and his Ph.D. in 1976, both from the University of Pittsburgh. His research interests are in the applications of probability and statistics to problems in reliability and quality control. His work in microelectronics reliability has been performed in collaboration with and under the support of the IBM Corp., INTELSAT, and the Bull Corp. He is the coauthor of over 50 refereed articles, has served in numerous editorial and referee capacities, and has lectured on reliability and maintenance topics throughout North America and Europe.

Dr. Nachlas' collaborative study of the deterioration of microelectronic communications circuits with J.L. Stevenson of INTELSAT earned them the 1991 P.K. McElroy Award. In addition, his work on the use of nested renewal functions to study opportunistic maintenance policies earned him the 2004 William A. Golomski Award.

Since 1986, Dr. Nachlas has served on the Management Committee for the Annual Reliability and Maintainability Symposium. Since 2002, he has been Proceedings Editor for the annual Reliability and Maintainability Symposium (RAMS). He is a member of INFORMS, IIE, and SRE, and is a Fellow of the ASQ.

During most of his tenure at Virginia Tech, Dr. Nachlas has also served as head coach of the men's lacrosse team, and in 2001 he was selected by U.S. Lacrosse as the men's division intercollegiate associates national coach of the year.

Contents

1

Introduction

Although we rarely think of it, reliability and maintenance are part of our everyday lives. The equipment, manufactured products, and fabricated infrastructure that contribute substantively to the quality of our lives have finite longevity. Most of us recognize this fact, but we do not always fully perceive the implications of finite system life for our efficiency and safety. Many, but not all, of us also appreciate the fact that our automobiles require regular service, but we do not generally think about the fact that roads and bridges, smoke alarms, electricity generation and transmission devices, and many others of the machines and facilities we use also require regular maintenance.

We are fortunate to live at a time in which advances in understanding of materials and energy have resulted in the creation of an enormous variety of sophisticated products and systems many of which (1) were inconceivable 100 or 200 or even 20 years ago, (2) contribute regularly to our comfort, health, happiness, efficiency, or success, (3) are relatively inexpensive, and (4) require little or no special training on our part. Naturally, our reliance on these devices and systems is continually increasing, and we rarely think about failure and the consequences of failure.

Occasionally, we observe a catastrophic failure. Fatigue failures of the fuselage of aircraft [1], the loss of an engine by a commercial jet [1], the Three Mile Island [1] and Chernobyl [1] nuclear reactor accidents, and the Challenger [2] and Discovery [3] space shuttle accidents are all widely known examples of catastrophic equipment failures. The relay circuit failure at the Ohio power plant that precipitated the August 2003 power blackout in the northeast United States and in eastern Canada [4] is an example of a system failure that directly affected millions of people. When these events occur, we are reminded dramatically of the fallibility of the physical systems on which we depend.

Nearly everyone has experienced less dramatic product failures such as that of a home appliance, the wearout of a battery, and failure of a light bulb. Many of us have also experienced potentially dangerous examples of product failures such as the blowout of an automobile tire.

Reliability engineering is the study of the longevity and the failure of equipment. Principles of science and mathematics are applied to the investigation of how devices age and fail. The intent is that a better understanding of device failure will aid in identifying ways in which product designs can be improved to increase life length and limit the adverse consequences of failure. The key point here is that the focus is upon design. New product and system designs must be shown to be safe and reliable prior to their fabrication and use. A dramatic example of a design for which the reliability was not properly evaluated is the well-known case of the Tacoma Narrows Bridge, which collapsed into the Puget Sound in November 1940, a few months after its completion [1].

The study of the reliability of an equipment design also has important economic implications for most products. As Blanchard [5] states, 90% of the life-cycle costs associated with the use of a product are fixed during the design phase of a product's life.

Similarly, an ability to anticipate failure can often imply the opportunity to plan for the efficient repair of equipment when it fails or, even better, to perform preventive maintenance in order to reduce failure frequency.

There are many examples of the products for which the system reliability is far better today than it was previously. One familiar example is the television set, which historically experienced frequent failures and which, at present, usually operates failure free beyond its age of obsolescence. Improved television reliability is certainly due largely to advances in circuit technology. However, the ability to evaluate the reliability of new material systems and of new circuit designs has also contributed to the gains we have experienced.

Perhaps the most well recognized system for which preventive maintenance is used to maintain product reliability is the commercial airplane. Regular inspection, testing, repair, and even overhaul are part of the normal operating life of every commercial aircraft. Clearly, the reason for such intense concern for the regular maintenance of aircraft is an appreciation of the influence of maintenance on failure probabilities and thus on safety.

On a personal level, the products for which we are most frequently responsible for maintenance are our automobiles. We are all aware of the inconvenience associated with an in-service failure of our cars and we are all aware of the relatively modest level of effort required to obtain the reduced failure probability that results from regular preventive maintenance.

It would be difficult to overstate the importance of maintenance and especially preventive maintenance. It is also difficult to overstate the extent to which maintenance is undervalued or even disliked. Historically, repair and especially preventive maintenance have often been viewed as inconvenient overhead activities that are costly and unproductive. Very rarely have the significant productivity benefits of preventive maintenance been recognized and appreciated. Recently, there are reports [6,7,8] that suggest that it is common experience for factory equipment to lose 10 to 40% of productive capacity to unscheduled repairs and that preventive maintenance could drastically reduce these losses. In fact, the potential productivity gains associated with the use of preventive maintenance strategies to reduce the frequency of unplanned failures constitute an important competitive opportunity [8]. The key to exploiting this opportunity is careful planning based on cost and reliability.

This book is devoted to the analytical portrayal and evaluation of equipment reliability and maintenance. As with all engineering disciplines, the language of description is mathematics. The text provides an exploration of the mathematical models that are used to portray, estimate, and evaluate device reliability and those that are used to describe, evaluate, and plan equipment service activities. In both cases, the focus is on design. The models of equipment reliability are the primary vehicle for recognizing deficiencies or opportunities to improve equipment designs. Similarly, using reliability as a basis, the models that describe equipment performance as a function of maintenance effort provide a means for selecting the most efficient and effective equipment service strategies.

The examples of various failures mentioned above share some common features, and they also have differences that are used here to delimit the extent of the analyses and discussions. Common features are that (1) product failure is sufficiently important that it warrants engineering effort to try to understand and control it, and (2) product design is complicated so the causes and consequences of failure are not obvious.

There are also some important differences among the examples. Taking an extreme case, the failure of a light bulb and the Three Mile Island reactor accident provide a defining contrast. The accident at Three Mile Island was precipitated by the failure of a physical component of the equipment. The progress and severity of the accident were also influenced by the response by humans to the component failure and by established decision policies. In contrast, the failure of a light bulb and its consequences are not usually intertwined with human decisions and performance. The point here is that there are very many modern products and systems for which operational performance depends upon the combined effectiveness of several of the following: (1) the physical equipment, (2) human operators, (3) software, and (4) management protocols.

It is both reasonable and prudent to attempt to include the evaluation of all four of these factors in the study of system behavior. However, the focus of this text is analytical, and the discussions are limited to the behavior of the physical equipment.

Several authors have defined analytical approaches to modeling the effects of humans [9] and of software [10] on system reliability. The motivation for doing this is the view that humans cause more system failures than does equipment. This view seems quite correct. Nevertheless, implementation of the existing mathematical models of human and software reliability requires the acceptance of the view that probability models appropriately represent dispersion in human behavior. In the case of software, existing models are based on the assumption that probability models effectively represent hypothesized evolution in software performance over time. The appropriateness of both of these points of view is subject to debate. It is considered here that the human operators of a system do not comprise a homogeneous population for which performance is appropriately modeled using a probability distribution. Similarly, software and operating protocols do not evolve in a manner that one would model using probability functions. As the focus of this text is the definition of representative probability models and their analysis, the discussion is limited to the physical devices.

The space shuttle accidents serve to motivate our focus on the physical behavior of equipment. The 1986 Challenger accident has been attributed to the use of the vehicle in an environment that was more extreme than the one for which it was designed. The 2002 Discovery accident is believed to have been the result of progressive deterioration at the site of damage to its heat shield. Thus, the physical design of the vehicles and the manner in which they were operated were incompatible, and it is the understanding of this interface that we obtain from reliability analysis.

The text is organized in four general sections. The early chapters describe in a stepwise manner the increasingly complete models of reliability and failure. These initial discussions include the key result that our understanding of design configurations usually implies that system reliability can usually be studied at the component level. This is followed by an examination of statistical methods for estimating reliability. A third section is comprised of five chapters that treat increasingly more complicated and more realistic models of equipment maintenance activities. Finally, several advanced topics are treated in the final chapter.

It is hoped that this sequence of discussions will provide the reader with a basis for further exploration of the topics treated. The development of new methods and models for reliability and maintenance has expanded our understanding significantly and is continuing. The importance of preventive maintenance for safety and industrial productivity is receiving increased attention. The literature that is comprised of reports of new ideas is expanding rapidly. This book is intended to prepare the reader to understand and use the new ideas as well as those that are included here.

As a starting point, note that it often happens that technical terms are created using words that already have colloquial meanings that do not correspond perfectly with their technical usage. This is true of the word reliability. In the colloquial sense, the word reliable is used to describe people who meet commitments. It is also used to describe equipment and other inanimate objects that operate satisfactorily. The concept is clear but not particularly precise. In contrast, for the investigations we undertake in this text, the word reliability has a precise technical definition. This definition is the departure point for our study.

2

System Structures

The point of departure for the study of reliability and maintenance planning is the elementary definition of the term reliability. As mentioned in Chapter 1, the technical definition of reliability is similar to the colloquial definition but is more precise. Formally, the definition is

> *Defn. 2.1*: *Reliability* is the probability that a device properly performs its intended function over time when operated within the environment for which it is designed.

Observe that there are four specific attributes of this definition of reliability. The four attributes are (1) probability, (2) proper performance, (3) qualification with respect to environment, and (4) time. All four are important. Over this and the next several chapters, we explore a series of algebraic models that are used to represent equipment reliability. We develop the models successively by sequentially including in the models each of the four attributes identified in the above definition. To start, consider the representation of equipment performance to which we refer as function.

2.1 STATUS FUNCTIONS

The question of what constitutes proper operation or proper function for a particular type of equipment is usually specific to the equipment. Rather than attempt to suggest a general definition for proper function, we assume that the appropriate definition for a device of interest has been specified, and we represent the functional status of the device as

$$\phi\begin{cases}1 & \textit{if the device functions properly}\\ 0 & \textit{if the device is failed}\end{cases}$$

Note that this representation is intentionally binary. We assume here that the status of the equipment of interest is either satisfactory or failed. There are many types of equipment for which one or more derated states are possible. Discussion of this possibility is postponed until the end of this chapter.

We presume that most equipment is comprised of components and that the status of the device is determined by the status of the components. Accordingly, let n be the number of components that make up the device and define the *component status variables*, x_i, as

$$x_i\begin{cases}1 & \textit{if component i is functioning}\\ 0 & \textit{if component i is failed}\end{cases}$$

so the set of n components that comprise a device is represented by the *component status vector*:

$$\underline{x} = \left\{x_1, x_2, \ldots\ldots.., x_n\right\}$$

Next, we represent the dependence of the device status on the component status as the function

$$\phi = \phi(\underline{x}) \qquad (2.1)$$

and the specific form for the function is determined by the way in which the components interact to determine system function. In the discussions that follow, $\phi(\underline{x})$ is referred to as a "system structure function" or as a "system status function" or simply as a "structure." In all cases, the intent is to reflect the dependence of the system state upon the states of the components that comprise the system. A parenthetical point is that the terms "device" and "system" are used here in a generic sense and may be interpreted as appropriate.

An observation concerning the component status vector is that it is defined here as a vector of binary elements so that an n-component system has 2^n possible component status vectors. For example, a three-component system has $2^3 = 8$ component status vectors. They are

$$\begin{array}{ll} \{1, 1, 1\} & \{1, 0, 1\} \\ \{1, 1, 0\} & \{1, 0, 0\} \\ \{0, 1, 1\} & \{0, 0, 1\} \\ \{0, 1, 0\} & \{0, 0, 0\} \end{array}$$

Each component status vector yields a corresponding value for the system status function, ϕ.

From a purely mathematical point of view, there is no reason to limit the definition of the system status function, so forms that have no practical interpretation can be constructed. In order to avoid any mathematically correct but practically meaningless forms for the system status function, we limit our attention to *coherent systems*.

> *Defn. 2.2*: A *coherent system* is one for which the system structure function is nondecreasing in each of its arguments.

This means that for each element of the component status vector, x_i, there exists a realization of the vector for which

$$\phi(x_1, ..., x_{i-1}, 0, x_{i+1}, ..., x_n) < \phi(x_1, ..., x_{i-1}, 1, x_{i+1}, ..., x_n) \quad (2.2)$$

Throughout our study of reliability, we will limit our attention to algebraic forms that comply with this restriction.

Generally, we expect that the physical relationships among the components determine the algebraic form of the system status function, ϕ.

2.2 SYSTEM STRUCTURES AND STATUS FUNCTIONS

Among reliability specialists, it is generally accepted that there are four generic types of structural relationships between a device and its components. These are (1) series, (2) parallel, (3) k out of n, and (4) all others. Consider each of these forms in sequence.

2.2.1 Series Systems

The simplest and most commonly encountered configuration of components is the series system. The formal definition of a series system is:

> *Defn. 2.3*: A *series system* is one in which all components must function properly in order for the system to function properly.

The conceptual analog to the series structure is a series-type electrical circuit. However, unlike a series circuit, it is specifically not implied here that the components must be physically connected in sequence. Instead, the point of emphasis is the requirement that all components function. An example of a series system in which the components are not physically connected is the set of legs of a three-legged stool. Another is the set of tires on an automobile. In both examples, the components are not physically connected to each other in a linear configuration. Nevertheless, all of the components must function properly for the system to operate.

The concept of a series circuit is commonly used to define a graphical representation of a series structure. For three components, this is shown in Figure 2.1. In general, representations of system structures such as the one in Figure 2.1 are referred to as reliability block diagrams. They are often helpful in understanding the relationships between components.

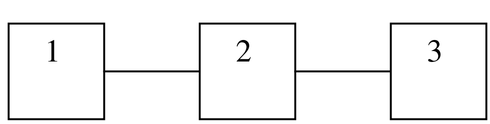


Figure 2.1 Reliability block diagram for a series system.

For the series structure, the requirement that all components must function in order for the system to function implies that a logical algebraic form for the system structure function is

$$\phi(\underline{x}) = \min_i \{x_i\} \tag{2.3}$$

but an equivalent and more useful form is

$$\phi(\underline{x}) = \prod_{i=1}^{n} x_i \tag{2.4}$$

As examples, consider a three-component series system and the cases

$$x_1 = x_2 = 1,\ x_3 = 0 \text{ and } \phi(\underline{x}) = 0$$

$$x_1 = 1,\ x_2 = x_3 = 0 \text{ and } \phi(\underline{x}) = 0$$

$$x_1 = x_2 = x_3 = 1 \text{ and } \phi(\underline{x}) = 1$$

Only the functioning of all components yields system function.

2.2.2 Parallel System

The second type of structure is the parallel structure. The conceptual analog is again the corresponding electrical circuit, and the definition is:

> *Defn. 2.4*: A *parallel system* is one in which the proper function of any one component implies system function.

It is again emphasized that no specific physical connections among the components are implied by the definition or by the reliability block diagram. Figure 2.2 shows the reliability block diagram for a three-component parallel system.

One example of a parallel system is the set of two engines on a two-engine airplane. As long as at least one engine functions, flight is sustained. However, this example implies that simply maintaining flight corresponds to proper function.

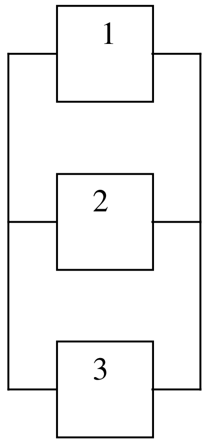


Figure 2.2 Reliability block diagram for a three-component parallel system.

It is a worthwhile debate to discuss when this is and when it is not an appropriate example of a parallel system.

Another example that is more appealing is the fact that the communications satellites presently in use have triple redundancy for each communications channel. That is, three copies of each set of transmitting components are installed in the satellite and arranged in parallel in order to assure continued operation of the channel. In view of the fact that this implies significant weight increases over the use of only single-configuration transmitters, the satellite provides an example of the importance of reliability as well as one of a parallel structure.

In a similar manner to that for the series system, the structure function for the parallel system may be defined as

$$\phi(\underline{x}) = \max_i \{x_i\} \tag{2.5}$$

An alternate form that is more amenable to analytical manipulation can be defined using a shorthand developed by Barlow and Proschan [11].

$$\phi(\underline{x}) = \coprod_{i=1}^{n} x_i \tag{2.6}$$

The inverted product symbol, $\coprod$, is called "ip" and is defined as

$$\coprod_{i=1}^{n} x_i = 1 - \prod_{i=1}^{n} (1 - x_i) \tag{2.7}$$

Once mastered, this shorthand is very convenient. Example cases for the three component parallel system are

$$x_1 = x_2 = 1,\ x_3 = 0, \text{ and } \phi(\underline{x}) = 1$$

$$x_1 = 1,\ x_2 = x_3 = 0, \text{ and } \phi(\underline{x}) = 1$$

$$x_1 = x_2 = x_3 = 0, \text{ and } \phi(\underline{x}) = 0$$

Conceptually, a parallel system is failed only when all system components are failed.

Before leaving the discussion of parallel structures, it is appropriate to mention the fact that the parallel arrangement of components is often referred to as *redundancy*. This is because the proper function of any of the parallel components implies proper function of the structure. Thus, the additional components are redundant until a component fails. Frequently, parallel structures are included in product designs specifically because of the resulting redundancy. Often but not always, the parallel components are identical. At the same time, there are actually several ways in which the redundancy may be implemented. A distinction is made between redundancy obtained using a parallel structure in which all components function simultaneously and that obtained using parallel components of which one functions and the other(s) wait as standby units until the failure of the functioning unit. Models that describe the reliability of active redundancy and of standby redundancy are presented at the end of Chapter 4.

2.2.3 k-out-of-n Systems

The third type of structure is the k-out-of-n structure. There is no obvious conceptual analog for this structure. A formal definition of it is:

> *Defn. 2.5*: A *k-out-of-n system* is one in which the proper function of any k of the n components that comprise the system implies proper system function.

The usual approach to constructing the reliability block diagram for the k-out-of-n system is to show a parallel diagram and to provide an additional indication that the system is k out of n.

An example of a k-out-of-n system is the rear axle of a large tractor-trailer on which the functioning of any three out of the four wheels is sufficient to assure mobility. Another example is the fact that some (1–k) electronic memory arrays are configured so that the operation of any 126 of the 128 memory addresses corresponds to satisfactory operation.

The algebraic representation of the structure function for a k-out-of-n system is not as compact as those for series and parallel systems. Given the definition of the relationship between component and system status, the most compact algebraic form for the structure function is

$$\phi(\underline{x}) = \begin{cases} 1 & if \quad \sum_{i=1}^{n} x_i \geq k \\ 0 & otherwise \end{cases} \tag{2.8}$$

Example cases for a 3 out of 4 system are

$$x_1 = x_2 = x_3 = 1,\ x_4 = 0, \text{ and } \phi(\underline{x}) = 1$$

$$x_1 = x_2 = 1,\ x_3 = x_4 = 0, \text{ and } \phi(\underline{x}) = 0$$

$$x_1 = x_2 = x_3 = 0,\ x_4 = 1, \text{ and } \phi(\underline{x}) = 0$$

Note that a series system may be considered as an n-out-of-n system and a parallel system may be viewed as a 1-out-of-n system. Thus, the k-out-of-n form provides a generalization that is sometimes useful in analyzing system performance in a generic context.

As indicated above, the fourth class of component configurations is the set of all others that are conceivable. This statement is not intended to be misleading. Instead, it is intended to imply that we can establish an equivalence between any arbitrary component configuration and one based on series and parallel structures. The process of constructing equivalent structures is explained in the next section and is illustrated with a classic example.

2.2.4 Equivalent Structures

The selection of a component configuration is usually made by the device designer in order to assure a specific functional capability. The configuration selected may not match one of the classes discussed above. In such a case, there are two ways by which we can obtain equivalent structural forms that may be easier to analyze than the actual one. The two ways are to use either minimum-path or minimum-cut analyses of the network representation of the system.

As a vehicle for illustrating the two methods, we use the Whetstone bridge. The reliability block diagram for the bridge is shown in Figure 2.3. Notice that the bridge structure is not

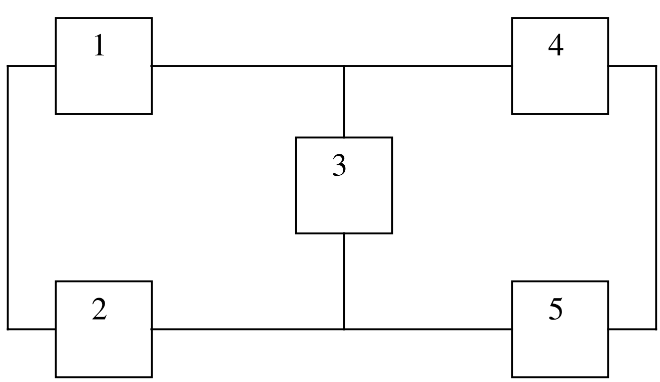


Figure 2.3 Reliability block diagram for a Whetstone bridge.

Table 2.1 System Status Values for the Bridge Structure

$\underline{x}$	$\phi(\underline{x})$	$\underline{x}$	$\phi(\underline{x})$
{1, 1, 1, 1, 1}	1	{0, 1, 1, 1, 1}	1
{1, 1, 1, 1, 0}	1	{0, 1, 1, 1, 0}	1
{1, 1, 1, 0, 1}	1	{0, 1, 1, 0, 1}	1
{1, 1, 1, 0, 0}	0	{0, 1, 1, 0, 0}	0
{1, 1, 0, 1, 1}	1	{0, 1, 0, 1, 1}	1
{1, 1, 0, 1, 0}	1	{0, 1, 0, 1, 0}	0
{1, 1, 0, 0, 1}	1	{0, 1, 0, 0, 1}	1
{1, 1, 0, 0, 0}	0	{0, 1, 0, 0, 0}	0
{1, 0, 1, 1, 1}	1	{0, 0, 1, 1, 1}	0
{1, 0, 1, 1, 0}	1	{0, 0, 1, 1, 0}	0
{1, 0, 1, 0, 1}	1	{0, 0, 1, 0, 1}	0
{1, 0, 1, 0, 0}	0	{0, 0, 1, 0, 0}	0
{1, 0, 0, 1, 1}	1	{0, 0, 0, 1, 1}	0
{1, 0, 0, 1, 0}	1	{0, 0, 0, 1, 0}	0
{1, 0, 0, 0, 1}	0	{0, 0, 0, 0, 1}	0
{1, 0, 0, 0, 0}	0	{0, 0, 0, 0, 0}	0

series, parallel, or k-out-of-n. Thus, the above algebraic representations cannot be used directly to provide a statement of the system status function.

We can obtain a system status function for the bridge in several ways. One obvious approach is to enumerate all of the component status vectors, to determine the system status for each vector, and to construct a table of system status values. For the five-component bridge structure, this is readily done, and the result is the given in Table 2.1. On the other hand, systems having a greater number of components cannot be handled so easily.

The use of minimum paths will permit us to analyze the bridge structure and other larger systems as well. Start with some definitions:

> *Defn. 2.6*: A *path vector*, $\underline{x}$, is a component status vector for which the corresponding system status function has a value of 1.

> *Defn. 2.7*: A *minimum-path vector*, $\underline{x}$, is a path vector for which any vector $\underline{y} < \underline{x}$ has a corresponding system status function with a value of 0.

Defn. 2.8: A *minimum-path set, P_j,* is the set of indices of a minimum-path vector for which the component status variable has a value of 1.

These definitions identify the component status vectors that correspond to system function and those vectors that are minimal in the sense that any reduction in the number of functioning components implies system failure. For the bridge structure, the minimum-path vectors and minimum-path sets are

$$\{1, 0, 0, 1, 0\} \quad P_1 = \{1, 4\}$$
$$\{0, 1, 0, 0, 1\} \quad P_2 = \{2, 5\}$$
$$\{1, 0, 1, 0, 1\} \quad P_3 = \{1, 3, 5\}$$
$$\{0, 1, 1, 1, 0\} \quad P_4 = \{2, 3, 4\}$$

Next, consider the elements of a minimum path and define a status function for each minimum path. That is, represent the functional status of each path using the functions $\rho(\underline{x})$. Since all of the components in a minimum path must function in order for the path to represent proper function, the components in a minimum path may be viewed as a series system. Hence, in general,

$$\rho_j(\underline{x}) = \prod_{i \in P_j} x_i \tag{2.9}$$

and for the example bridge structure,

$$\rho_1(\underline{x}) = \prod_{i \in P_1} x_i = x_1 x_4$$

$$\rho_2(\underline{x}) = \prod_{i \in P_2} x_i = x_2 x_5$$

$$\rho_3(\underline{x}) = \prod_{i \in P_3} x_i = x_1 x_3 x_5$$

$$\rho_4(\underline{x}) = \prod_{i \in P_4} x_i = x_2 x_3 x_4$$

Now, observe that the original system will function if any of the minimum paths is functioning. Therefore, we may view the system as a parallel arrangement of the minimum paths. Algebraically, this means

$$\phi(\underline{x}) = \coprod_j \rho_j(\underline{x}) = \coprod_j \prod_{i \in P_j} x_i \tag{2.10}$$

For the bridge structure, this expression expands to

$$\phi(\underline{x}) = 1 - (1 - x_1 x_4)(1 - x_2 x_5)(1 - x_1 x_3 x_5)(1 - x_2 x_3 x_4)$$

The most important point here is that, for any component status vector, the Expression 2.10 will always give the same system status value as Table 2.1. That is, the parallel arrangement of the minimum paths of a system with the components of the respective minimum paths arranged in series constitutes a system that is equivalent to the original system. Figure 2.4 shows the graphical realization of this equivalence for the bridge structure.

It is appropriate to emphasize here the fact that the equivalent structure has exactly the same status function value as the original structure for all realizations of the component status vector. Consequently, the minimum-path analysis permits us to identify a form for the system status function that can be computed using only series and parallel algebraic forms.

There is a comparable construction using the idea of cut vectors rather than path vectors. The method based on cut vectors may also be used for the bridge and other structures. Again, start with some definitions:

> *Defn. 2.9*: A *cut vector*, $\underline{x}$, is a component status vector for which the corresponding system status function has a value of 0.

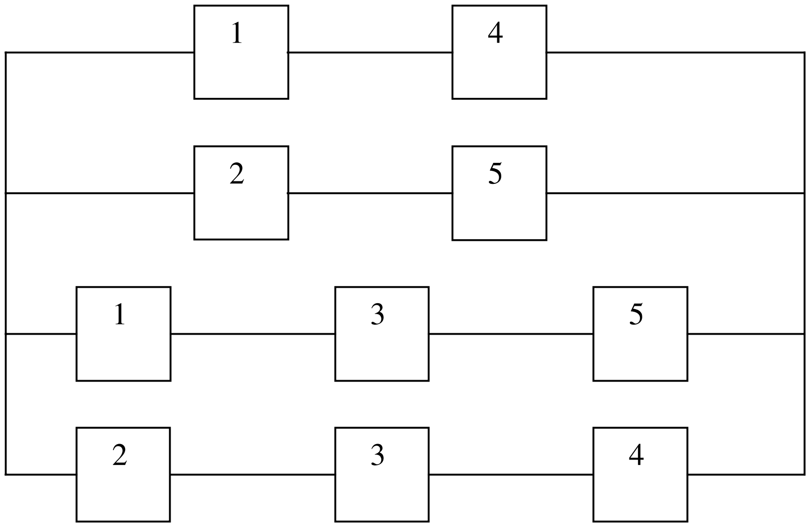


Figure 2.4 Minimum path equivalent structure for the Whetstone bridge.

Defn. 2.10: A *minimum-cut vector*, $\underline{x}$, is a cut vector for which any vector $\underline{y} > \underline{x}$ has a corresponding system status function with a value of 1.

Defn. 2.11: A *minimum-cut set, C_k*, is the set of indices of a minimum-cut vector for which the component status variable has a value of 0.

These definitions identify the component status vectors that correspond to system failure and those vectors that are minimal in the sense that any increase in the number of functioning components implies satisfactory system function. For the bridge structure, the minimum-cut vectors and minimum-cut sets are

$$\{0, 0, 1, 1, 1\} \qquad C_1 = \{1, 2\}$$

$$\{1, 1, 1, 0, 0\} \qquad C_2 = \{4, 5\}$$

$$\{0, 1, 0, 1, 0\} \qquad C_3 = \{1, 3, 5\}$$

$$\{1, 0, 0, 0, 1\} \qquad C_4 = \{2, 3, 4\}$$

Based on the definition of the minimum cuts, we see that the system will function if any of the elements of the minimum-cut function. Hence, we may define a structure function for the minimum cuts as

$$\kappa_k(\underline{x}) = \coprod_{i \in C_k} x_i \tag{2.11}$$

In general and for the specific case of the bridge structure, we have

$$\kappa_1(\underline{x}) = \coprod_{i \in C_1} x_i = 1 - (1 - x_1)(1 - x_2)$$

$$\kappa_2(\underline{x}) = \coprod_{i \in C_2} x_i = 1 - (1 - x_4)(1 - x_5)$$

$$\kappa_3(\underline{x}) = \coprod_{i \in C_3} x_i = 1 - (1 - x_1)(1 - x_3)(1 - x_5)$$

$$\kappa_4(\underline{x}) = \coprod_{i \in C_4} x_i = 1 - (1 - x_2)(1 - x_3)(1 - x_4)$$

We observe further that the system will function only if all of the minimum cuts are inactive — if all are functioning. If any minimum cut is active, the system is failed, so the minimum cuts act as a series system with respect to system operation (Figure 2.5).

Here again, it is appropriate to emphasize the fact that the equivalent structure and the original structure have the same status function value for each component status vector. Thus, the system status may be calculated using only the simple series and parallel forms.

One further observation concerning the equivalent structures is that one may use either the minimum-cut or the minimum-path method. Both yield equivalent expressions for the system status so we may use the one that appears easier or preferable for some other reason.

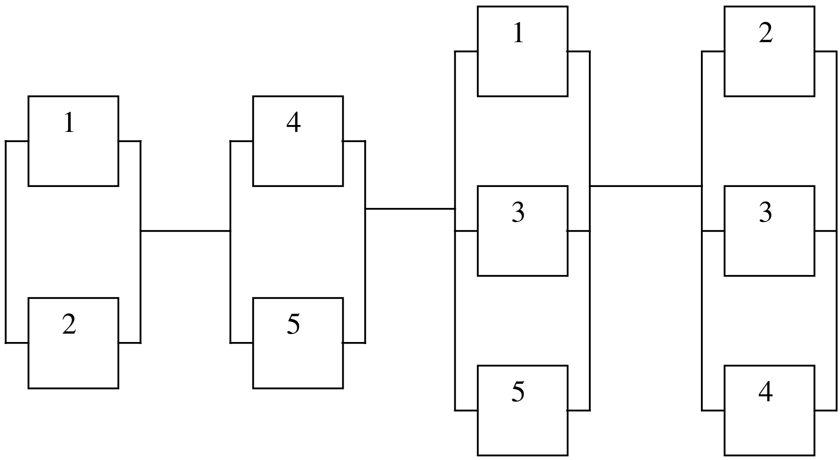


Figure 2.5 Minimum cut equivalent structure for the Whetstone bridge.

2.3 MODULES OF SYSTEMS

Most modern equipment is fairly complicated and is comprised of relatively many components. For example, depending upon how one counts them, one might say that there are about 100 components in a television, 300 components in a personal computer, and 600 components in a medium-size automobile. Even with minimum paths or minimum cuts, evaluating the status function for systems of this scale is too difficult. However, it is also common to find that a complicated system such as a television is actually comprised of subsystems called modules. A television usually has a power management module, a video signal reception module, an audio signal reception module, a sound production module, a video projection module, and some sort of system control module. Thus, the system may be viewed as comprised of "super components" called modules. System status may be defined as a function of the status modules, and the status of each module should be a function of the components that comprise it.

To formalize this idea algebraically, assume that the n components that comprise a system can be partitioned into m sets of components corresponding to m modules. Let $\psi_l(\underline{x})$ represent the module status function for module l, where l = 1, … , m. Then

$$\underline{\psi}(\underline{x}) = \{\psi_1(\underline{x}), \psi_2(\underline{x}), \ldots\ldots, \psi_m(\underline{x})\}$$

is a vector of binary module status values, and the system status is defined as

$$\phi(\underline{x}) = \phi(\underline{\psi}(\underline{x})) \tag{2.12}$$

Naturally, the module status functions may be evaluated using minimum paths or cuts, and the system status may be analyzed by treating the modules as components and applying the minimum-path or minimum-cut methods as appropriate.

Of course, the partition of the set of components may be performed at several levels if that is appropriate. The principle remains the same. Any meaningful decomposition of the system components may be used to simplify the representation and analysis of the system status.

A final point here is that systems generally have only one of each module and usually require that all modules function properly in order for the system to operate satisfactorily. Thus, while it is not always the case, the modules often comprise a series structure. Consequently, the analysis of system behavior in terms of modules can be quite efficient.

2.4 MULTISTATE COMPONENTS AND SYSTEMS

The treatment of component and system status throughout most of this text is limited to the case in which system state and component states are binary. Nevertheless, it should be recognized that for some equipment multiple states may be meaningful. It is reasonable to define derated or otherwise incomplete levels of performance for some equipment. It may even be appropriate for some devices to define a continuous state variable on the interval [0, 1]. Algebraic models for system state using multistate components and for multistate systems with binary components have been studied. The key references for this work are the papers by Natvig [12] and by Barlow and Wu [13]. In each case, the key to the construction is the algebraic representation of the effect of component state on system state.

Both Natvig and Barlow and Wu start by defining the state space for the components as {0, 1, … , m} where $x_i = j$ represents the condition that component i is in state j, the state 0 corresponds to component failure, and the state m corresponds to perfect functioning. The interpretation of the intermediate states depends upon the physical characteristics of the specific component. Using this basic format, there are two immediately apparent approaches to defining the system state as a function of the states of the components of a system.

One approach is to define a binary system state by partitioning all of the possible component status vectors into two subsets, one for which system performance is acceptable and the other for the case that system level performance is unacceptable. Naturally, this cannot be done in general, because the selection of members of the two subsets depends upon the particular system.

An alternate, and more general, approach is to define the system state to also be an element of the set {0, 1, … , m} and to define the value of the system state using minimum-path and minimum-cut concepts. To do this, we first specify that, as in the case of the binary state space, the state of a series system of multistate components is the minimum of the component state values. Thus, Equation 2.3 applies:

$$\phi(\underline{x}) = \min_i \left\{x_i\right\} \tag{2.13}$$

In the same manner, the state of a parallel system is the maximum of the component state values as stated in Equation 2.5:

$$\phi(\underline{x}) = \max_i \left\{x_i\right\} \tag{2.14}$$

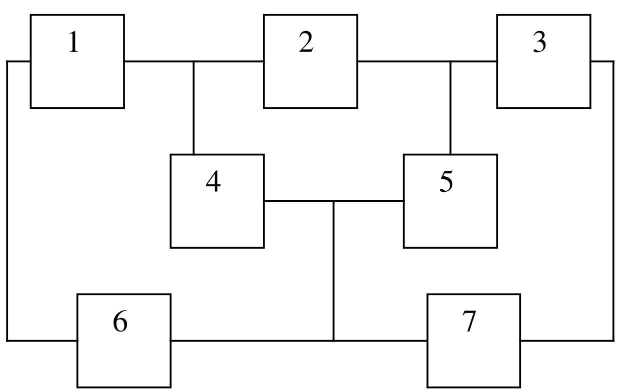
Then for more general structures, we use the min paths or min cuts to define

$$\phi(\underline{x}) = \max_j \left\{ \min_{i \in P_j} \left\{ x_i \right\} \right\} = \min_k \left\{ \max_{i \in C_k} \left\{ x_i \right\} \right\} \tag{2.15}$$

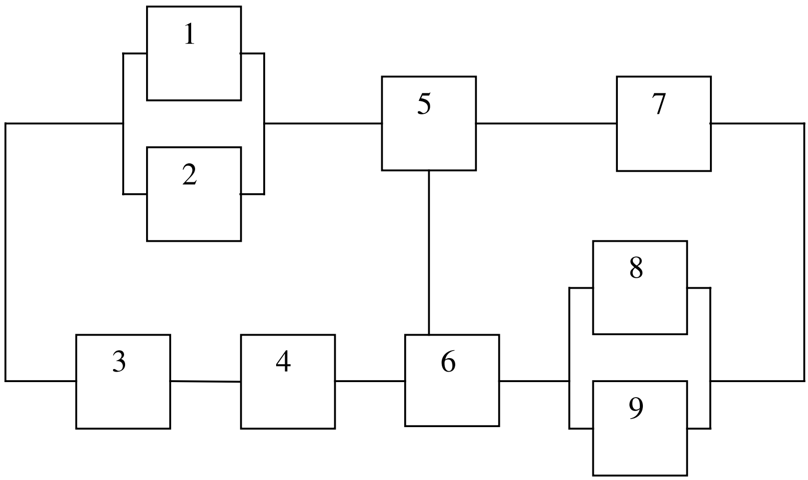
This general formulation may be tailored to nearly any application. For example, one can restrict some of the components to only a subset of the $m+1$ states. One may also incorporate the multistate status measure within a modular decomposition in whatever manner is meaningful. Finally, this formulation has the appealing feature that it subsumes the binary case.

2.5 EXERCISES

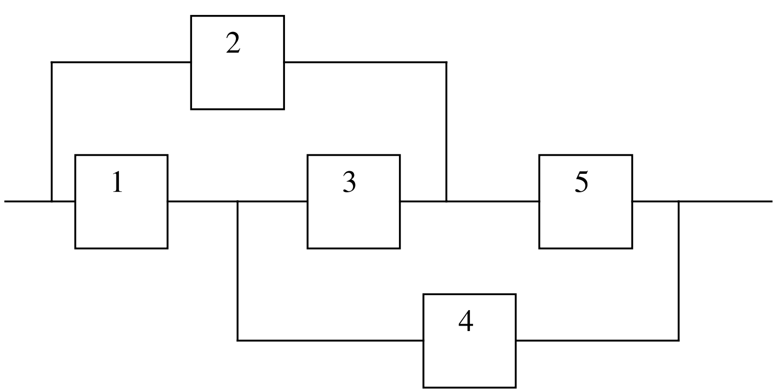
1. Construct the minimum-path and minimum-cut equivalent structures for a 2-out-of-3 system.
2. Construct the minimum-path and minimum-cut equivalent structures for a three-component series system.
3. Construct the minimum-path and minimum-cut equivalent structures for a three-component parallel system.
4. Identify the primary modules of an automobile.
5. Construct the minimum-path and minimum-cut equivalent structures for the following system:



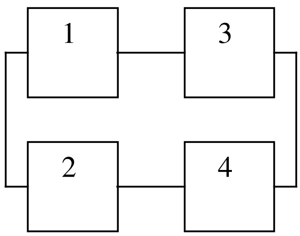
6. Construct the minimum-path and minimum-cut equivalent structures for the following system:

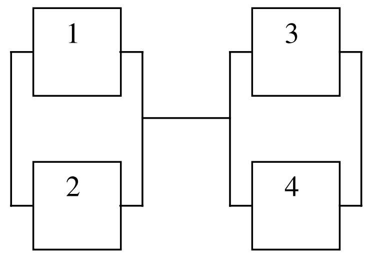


7. Construct the minimum-path and minimum-cut equivalent structures for the following system:



8. Construct and compare the minimum-path and minimum-cut equivalent structures for the following systems:





3

Reliability of System Structures

The next logical step in our construction of mathematical models of reliability is to enhance the system structure models by the addition of probability — the second of the attributes of the definition of reliability. As we do this, we will refer to the probabilities as reliabilities despite the fact that we have not yet included all four attributes in our models. Naturally, we expect that the reliability of a system will be represented as a function of the reliabilities of its constituent components.

3.1 PROBABILITY ELEMENTS

Keeping in mind the fact that we represent system state as a binary variable, ϕ, define the system reliability, R_s, to be the probability that the system is functioning:

$$R_s = \Pr[\phi = 1] \tag{3.1}$$

Observe that an artifact of the binary definition of the system state is that the system reliability is also the expected value of the system state variable:

$$E[\phi] = 1 \cdot \Pr[\phi = 1] + 0 \cdot \mathrm{P}[\phi = 0] = \Pr[\phi = 1] \tag{3.2}$$

A similar pair of definitions applies to the component status variables. That is, we let

$$r_i = \Pr[x_i = 1] \tag{3.3}$$

where, because of the fact that the x_i are binary, it is again the case that the reliability and expected value correspond. For a system comprised of n components, we take

$$\underline{r} = \{r_1, r_2, \ldots\ldots, r_n\}$$

to be the vector of component reliability values. Given the defined notation, it is reasonable to expect that the system reliability can be expressed as a function of the component reliabilities. In terms of general notation, we represent this as

$$R_s(\underline{r}) = \Pr[\phi(\underline{x}) = 1] \tag{3.4}$$

and we devote the next section to the realizations of this expression.

3.2 RELIABILITY OF SYSTEM STRUCTURES

The formulation of the system reliability function is often relatively straightforward, but the general forms must be constructed carefully. The key issue to consider is whether or not the components are mutually independent. In this construction, we follow the same four cases that we examined in the previous chapter.

3.2.1 Series Systems

Based on the form of the system structure function, the general statement of the reliability function for a series system is

$$R_s = \Pr[\phi(\underline{x}) = 1] = \Pr[\prod_{i=1}^{n} x_i = 1] \tag{3.5}$$

Now, in general

$$\Pr[\prod_{i=1}^{n} x_i = 1] \geq \prod_{i=1}^{n} \Pr[x_i = 1]$$

and equality holds only when the components are mutually independent. Thus, for a series system comprised of independent components, the system reliability may be stated as

$$R_s = \prod_{i=1}^{n} r_i \tag{3.6}$$

We should note that, regardless of whether or not the components are independent, the system reliability function is an increasing function of the component reliability values and is a decreasing function of the number of components.

In most series systems, the components are independent with respect to their probabilities of proper function. One noteworthy class of structures for which the components are not independent is the set of systems for which the components share loads. This is discussed further in Chapter 14.

3.2.2 Parallel Systems

The expression corresponding to Equation 3.5 for a parallel system is

$$R_s = \Pr[\phi(\underline{x}) = 1] = \Pr[\coprod_{i=1}^{n} x_i = 1] = \Pr[\prod_{i=1}^{n} (1 - x_i) = 0] \tag{3.7}$$

and it is again the case that

$$\Pr[\coprod_{i=1}^{n} x_i = 1] \le \coprod_{i=1}^{n} \Pr[x_i = 1]$$

so a parallel system of independent components has reliability function

$$R_s = \coprod_{i=1}^{n} r_i \tag{3.8}$$

Examination of this function indicates that the system reliability function for a parallel system is increasing in both the component reliability values and in the number of components.

For system structures other than series and parallel, the computation of the system reliability from the component reliabilities is not as straightforward.

3.2.3 k-out-of-n Systems

For the generic k-out-of-n system, there is no compact statement of the reliability function. We can only state that

$$R_s = \Pr[\sum_{i=1}^{n} x_i \geq k] \tag{3.9}$$

and even when the components are independent, there is no convenient form for this function. The single exception occurs when the n components are independent and identical (have the same reliability). In that case, the system reliability is given by the sum of binomial probabilities for k or more functioning components:

$$R_s = \sum_{j=k}^{n} \binom{n}{j} r^j (1-r)^{n-j} \tag{3.10}$$

For most other k-out-of-n systems, the use of the minimum-path- and minimum-cut-based methods of the next section provide the most effective approach to evaluating system reliability.

3.2.4 Equivalent Structures

The fact that every system structure has an equivalent representation as a combination of series and parallel forms suggests that we can use the equivalent structures to evaluate system reliability. While this is true, the fact that the minimum paths and minimum cuts that comprise the equivalent forms are usually not independent implies that we will be able to obtain bounds on system reliability rather than exact system reliability values. The process of using the minimum paths and minimum cuts is again illustrated by the Whetstone bridge shown in Figure 2.3

Before developing the minimum-path- and minimum-cut-based bounds, let us note that, using the same approach as we did for the system status, we can determine the system reliability by complete enumeration. Specifically, we can enumerate all possible system states, obtain the probability that each state occurs, and sum those probabilities that correspond to system function. For the case in which the components are independent, the result of the enumeration is shown in Table 3.1.

Based on the expressions in Table 3.1, the system reliability is

$$
\begin{aligned}
R_s = {} & r_1r_2r_3r_4r_5 + r_1r_2r_3r_4(1 - r_5) + r_1r_2r_3(1 - r_4)r_5 \\
& + r_1r_2r(1 - r_3)r_4r_5 + r_1r_2(1 - r_3)r_4(1 - r_5) \\
& + r_1r_2(1 - r_3)(1 - r_4)r_5 + r_1(1 - r_2)r_3r_4r_5 \\
& + r_1(1 - r_2)r_3r_4(1 - r_5) + r_1(1 - r_2)r_3(1 - r_4)r_5 \\
& + r_1(1 - r_2)(1 - r_3)r_4r_5 + r_1(1 - r_2)(1 - r_3)r_4(1 - r_5) \\
& + (1 - r_1)r_2r_3r_4r_5 + (1 - r_1)r_2r_3r_4(1 - r_5) \\
& + (1 - r_1)r_2r_3(1 - r_4)r_5 + (1 - r_1)r_2(1 - r_3)r_4r_5 \\
& + (1 - r_1)r_2(1 - r_3)(1 - r_4)r_5 (3.11)
\end{aligned}
$$

With considerable algebraic effort, this reduces to

$$
\begin{aligned}
R_s = {} & r_1r_4 + r_2r_5 + r_1r_3r_5 + r_2r_3r_4 - r_1r_2r_3r_4 - r_1r_2r_3r_5 - r_1r_2r_4r_5 \\
& - r_1r_3r_4r_5 - r_2r_3r_4r_5 + 2r_1r_2r_3r_4r_5 \qquad (3.12)
\end{aligned}
$$

In addition, for a structure such as the bridge, it would be reasonable for all five components to be identical and to have the same reliability. Then,

$$r_i = r \qquad \forall i$$

and the system reliability function reduces to the polynomial

$$R_s = 2\,r^2 + 3r^3 - 5r^4 + 2r^5 \qquad (3.13)$$

Table 3.1 Reliability Values for Paths of the Whetstone Bridge

$\underline{x}$	$\phi(\underline{x})$	$\Pr[\phi(\underline{x}) = 1]$	$\underline{x}$	$\phi(\underline{x})$	$\Pr[\phi(\underline{x}) = 1]$
{1, 1, 1, 1, 1}	1	$r_1r_2r_3r_4r_5$	{0, 1, 1, 1, 1}	1	$(1 - r_1)r_2r_3r_4r_5$
{1, 1, 1, 1, 0}	1	$r_1r_2r_3r_4(1 - r_5)$	{0, 1, 1, 1, 0}	1	$(1 - r_1)r_2r_3r_4(1 - r_5)$
{1, 1, 1, 0, 1}	1	$r_1r_2r_3(1 - r_4)r_5$	{0, 1, 1, 0, 1}	1	$(1 - r_1)r_2r_3(1 - r_4)r_5$
{1, 1, 1, 0, 0}	0	$r_1r_2r_3r(1 - r_4)(1 - r_5)$	{0, 1, 1, 0, 0}	0	$(1 - r_1)r_2r_3r(1 - r_4)(1 - r_5)$
{1, 1, 0, 1, 1}	1	$r_1r_2(1 - r_3)r_4r_5$	{0, 1, 0, 1, 1}	1	$(1 - r_1)r_2(1 - r_3)r_4r_5$
{1, 1, 0, 1, 0}	1	$r_1r_2(1 - r_3)r_4(1 - r_5)$	{0, 1, 0, 1, 0}	0	$(1 - r_1)r_2(1 - r_3)r_4(1 - r_5)$
{1, 1, 0, 0, 1}	1	$r_1r_2(1 - r_3)(1 - r_4)r_5$	{0, 1, 0, 0, 1}	1	$(1 - r_1)r_2(1 - r_3)(1 - r_4)r_5$
{1, 1, 0, 0, 0}	0	$r_1r_2(1 - r_3)(1 - r_4)(1 - r_5)$	{0, 1, 0, 0, 0}	0	$(1 - r_1)r_2(1 - r_3)(1 - r_4)(1 - r_5)$
{1, 0, 1, 1, 1}	1	$r_1(1 - r_2)r_3r_4r_5$	{0, 0, 1, 1, 1}	0	$(1 - r_1)(1 - r_2)r_3r_4r_5$
{1, 0, 1, 1, 0}	1	$r_1(1 - r_2)r_3r_4(1 - r_5)$	{0, 0, 1, 1, 0}	0	$(1 - r_1)(1 - r_2)r_3r_4(1 - r_5)$
{1, 0, 1, 0, 1}	1	$r_1(1 - r_2)r_3(1 - r_4)r_5$	{0, 0, 1, 0, 1}	0	$(1 - r_1)(1 - r_2)r_3(1 - r_4)r_5$
{1, 0, 1, 0, 0}	0	$r_1(1 - r_2)r_3(1 - r_4)(1 - r_5)$	{0, 0, 1, 0, 0}	0	$(1 - r_1)(1 - r_2)r_3(1 - r_4)\ (1 - r_5)$
{1, 0, 0, 1, 1}	1	$r_1(1 - r_2)(1 - r_3)r_4r_5$	{0, 0, 0, 1, 1}	0	$(1 - r_1)(1 - r_2)(1 - r_3)r_4r_5$
{1, 0, 0, 1, 0}	1	$r_1(1 - r_2)\ (1 - r_3)r_4(1 - r_5)$	{0, 0, 0, 1, 0}	0	$(1 - r_1)(1 - r_2)(1 - r_3)r_4(1 - r_5)$
{1, 0, 0, 0, 1}	0	$r_1(1 - r_2)\ (1 - r_3)\ (1 - r_4)r_5$	{0, 0, 0, 0, 1}	0	$(1 - r_1)(1 - r_2)(1 - r_3)(1 - r_4)r_5$
{1, 0, 0, 0, 0}	0	$r_1(1 - r_2)(1 - r_3)(1 - r_4)(1 - r_5)$	{0, 0, 0, 0, 0}	0	$(1 - r_1)(1 - r_2)(1 - r_3)(1 - r_4)(1 - r_5)$

Clearly, system reliability evaluation by this method is rather demanding and offers plenty of possibility for error.

An alternate approach is to construct bounds on system reliability. Three sets of bounds on system reliability have been defined. It is appropriate to compute all three sets and to combine the information they provide to obtain the narrowest possible interval within which the system reliability will lie.

The first set of bounds is reasonably obvious and is generally not very tight. These are the series and parallel bounds. To compute these bounds, we simply treat the system components as if the system configuration were a series structure of independent components and calculate a lower bound, and we then calculate an upper bound assuming a parallel configuration. Thus,

$$b_s = \prod_{i=1}^{n} r_i \leq R_s \leq \coprod_{i=1}^{n} r_i = b_p \tag{3.14}$$

For a system such as the bridge structure, assuming all of the components are identical implies that these bounds reduce to

$$b_s = r^n \leq R_s \leq 1-(1-r)^n = b_p$$

Example calculations of these bounds are presented in Table 3.2.

A second set of reliability bounds may be constructed using minimum paths and minimum cuts. Remember that the minimum-cut equivalent structure has the minimum cuts arranged in series, and recall that for a series structure in general,

$$\Pr\left[\prod_{i=1}^{n} x_i = 1\right] \geq \prod_{i=1}^{n} \Pr\left[x_i = 1\right]$$

Applying this inequality to the minimum-path structures yields the minimum-cut lower bound on system reliability:

$$b_{mcl} = \prod_k \Pr\left[\kappa_k(\underline{x}) = 1\right] \le \Pr\left[\prod_k \kappa_k(\underline{x}) = 1\right]$$
$$= \Pr\left[\phi(\underline{x}) = 1\right] = R_s \tag{3.15}$$

The same reasoning can be applied to the minimum paths. The minimum-path equivalent structure has the minimum paths arranged in a parallel configuration. Therefore, applying the inequality

$$\Pr\left[\coprod_{i=1}^{n} x_i = 1\right] \le \coprod_{i=1}^{n} \Pr\left[x_i = 1\right]$$

to the minimum paths yields the minimum path upper bound on system reliability:

$$b_{mpu} = \coprod_j \Pr\left[\rho_j(\underline{x}) = 1\right] \ge \Pr\left[\coprod_j \rho_j(\underline{x}) = 1\right]$$
$$= \Pr\left[\phi(\underline{x}) = 1\right] = R_s \tag{3.16}$$

For the example bridge structure having five identical components, these bounds are computed as follows:

$$\kappa_1(\underline{x}) = \coprod_{i \in C_1} x_i = 1 - (1 - x_1)(1 - x_2) \ ,$$

so $$\Pr\left[\kappa_1(\underline{x}) = 1\right] = \coprod_{i \in C_1} r_i = 1 - (1 - r)^2$$

$$\kappa_2(\underline{x}) = \coprod_{i \in C_2} x_i = 1 - (1 - x_4)(1 - x_5) \ ,$$

so $$\Pr[\kappa_2(\underline{x}) = 1] = \coprod_{i \in C_2} r_i = 1 - (1 - r)^2$$

$$\kappa_3(\underline{x}) = \coprod_{i \in C_3} x_i = 1 - (1 - x_1)(1 - x_3)(1 - x_5) \ ,$$

so $$\Pr[\kappa_3(\underline{x})=1]=\coprod_{i\in C_3} x_i = 1-(1-r)^3$$

$$\kappa_4(\underline{x})=\coprod_{i\in C_4} x_i = 1-(1-x_2)(1-x_3)(1-x_4)\ ,$$

so $$\Pr[\kappa_4(\underline{x})=1]=\coprod_{i\in C_4} x_i = 1-(1-r)^3$$

$$\rho_1(\underline{x})=\prod_{i\in P_1} x_i = x_1x_4\ ,$$

so $$\Pr[\rho_1(\underline{x})=1]=\prod_{i\in P_1} r_i = r^2$$

$$\rho_2(\underline{x})=\prod_{i\in P_2} x_i = x_2x_5\ ,$$

so $$\Pr[\rho_2(\underline{x})=1]=\prod_{i\in P_2} r_i = r^2$$

$$\rho_3(\underline{x})=\prod_{i\in P_3} x_i = x_1x_3x_5\ ,$$

so $$\Pr[\rho_3(\underline{x})=1]=\prod_{i\in P_3} r_i = r^3$$

$$\rho_4(\underline{x})=\prod_{i\in P_4} x_i = x_2x_3x_4\ ,$$

so $$\Pr[\rho_4(\underline{x})=1]=\prod_{i\in P_4} r_i = r^3$$

Then,

$$b_{mcl} = \left(1-(1-r)^2\right)^2 \left(1-(1-r)^3\right)^2 \leq R_s \leq \left(1-(1-r^2)^2(1-r^3)^2\right)$$
$$= b_{mpu}$$

Example calculations of these bounds are also included in Table 3.2.

The minimum paths and minimum cuts may be used to define a third set of bounds. These are known as the minimax bounds. Starting with the minimum paths, recall that the structure function for a parallel system may also be expressed in terms of a maximum. That is, as stated in Equation 2.5,

$$\phi(\underline{x}) = \max_i \left\{x_i\right\}$$

so

$$\Pr[\phi(\underline{x}) = 1] = \Pr[\max_i \left\{x_i\right\} = 1]$$

and for a parallel arrangement of minimum paths,

$$\Pr[\phi(\underline{x}) = 1] = \Pr[\max_j \left\{\rho_j(\underline{x})\right\} = 1]$$

In general, for any set of probabilities

$$\max_j \left\{\Pr[\rho_j(\underline{x}) = 1]\right\} \leq \Pr[\max_j \left\{\rho_j(\underline{x})\right\} = 1]$$

so the minimax lower bound on system reliability is

$$b_{mml} = \max_j \left\{\Pr[\rho_j(\underline{x}) = 1]\right\} \leq R_s \tag{3.17}$$

Applying the same logic to the minimum cuts, we have from Equation 2.3

$$\phi(\underline{x}) = \min_i \left\{x_i\right\}$$

so

$$\Pr[\phi(\underline{x})=1]=\Pr[\min_i\{x_i\}=1]$$

and for the parallel configuration of the minimum cuts

$$\Pr[\phi(\underline{x})=1]=\Pr[\min_k\{\kappa_k\}=1]$$

In general,

$$\Pr[\min_k\{\kappa_k(\underline{x})\}=1]\le\min_k\{\Pr[\kappa_k(\underline{x})=1\}$$

And the resulting minimax upper bound on system reliability is

$$R_s\le\min_k\{\Pr[\kappa_k(\underline{x})=1\}=b_{mmu} \tag{3.18}$$

Therefore, the minimax reliability bounds are

$$b_{mml}=\max_j\{\Pr[\rho_j(\underline{x})=1]\}\le R_s\le\min_k\{\Pr[\kappa_k(\underline{x})=1\}=b_{mmu} \tag{3.19}$$

For the bridge with identical components, the computation of the minimax bounds proceeds as follows:

$$b_{mml}=\max_j\{\Pr[\rho_j(\underline{x})=1]\}=\max\{r^2,r^2,r^3,r^3\}=r^2$$

and

$$\begin{aligned}b_{mmu}&=\min_k\{\Pr[\kappa_k(\underline{x})=1\}\\&=\min\{1-(1-r)^2,1-(1-r)^2,1-(1-r)^3,1-(1-r)^3\}\\&=1-(1-r)^2\end{aligned}$$

So

$$r^2\le R_s\le 1-(1-r)^2$$

To illustrate the computation and behavior of the bounds, we have calculated the values of each of the bounds for the bridge with identical components for several values of the

Table 3.2 Computed Values of the System Reliability and the Bounds on System Reliability

r	b_s	b_{mcl}	b_{mml}	R_s	b_{mmu}	b_{mpu}	b_p
0.99	0.951	0.999	0.980	0.999	1.000	1.000	1.000
0.95	0.774	0.995	0.903	0.995	0.998	1.000	1.000
0.90	0.591	0.978	0.810	0.979	0.990	0.997	1.000
0.75	0.273	0.852	0.563	0.861	0.938	0.936	0.999
0.60	0.078	0.618	0.360	0.660	0.840	0.748	0.990
0.50	0.031	0.431	0.250	0.500	0.750	0.569	0.969
0.25	0.001	0.064	0.063	0.139	0.438	0.148	0.763
0.10	0.000	0.003	0.010	0.022	0.190	0.022	0.410

component reliability. These values as well as the actual system reliability value (from Equation 3.13) are shown in Table 3.2.

Note that one may not in general assume that one of the sets of bounds will be tighter than the others. A reasonable approach to using the bounds is to take the greatest of the lower bounds and the smallest of the upper bounds to obtain the tightest possible bounds on system reliability. Doing this for the entries of Table 3.2 indicates that very satisfactory bounds are obtained.

3.3 MODULES

The idea that the components that comprise a system may sometimes be partitioned into modules may be extended to the calculation of system reliability bounds. There are three key ways in which this may be pursued. Recall that the algebraic representation of system state using modules is

$$\phi(\underline{x}) = \phi(\underline{\psi}(\underline{x}))$$

where

$$\underline{\psi}(\underline{x}) = \{\psi_1(\underline{x}), \psi_2(\underline{x}), \text{........}, \psi_m(\underline{x})\}$$

One possible approach to the use of the partition is to calculate the reliability of each module and to use the resulting values as component values in the bounds defined in the previous section. If the modules contain relatively few components or are configured in either a series of parallel structure, this method will be fairly straightforward.

If, on the other hand, some of the modules are themselves rather complicated, but the system is designed with the modules in series, then the series computation applied to each of the upper and the lower bounds provides a pair of bounds on system reliability. In this case, the bounds on module reliability are obtained using the methods of the previous section.

The third possibility is that one or more of the modules are complicated, and the system configuration of the modules is not a simple one. In this case, a lower bound on system reliability can be computed by applying the minimum-cut lower-bound calculation at the system level to the minimum-cut lower bounds for the modules.

3.4 RELIABILITY IMPORTANCE

In view of the influence of the reliabilities of the components, it is reasonable to ask which components have the greatest (or least) impact on system reliability. One possible reason for examining this question is to help decide which of the components should be improved first. The idea that the effect of each component is worth considering has led several authors to define various measures of "reliability importance." Among the several forms that have been suggested, the one based on derivatives seems to us to be the most logical. The formal definition is

> *Defn. 3.1*: The *reliability importance* of component i, $I_R(i)$, of a coherent system is the derivative of the system reliability function with respect to the component i reliability. That is,

$$I_R(i) = \frac{d}{dr_i} R_s \tag{3.20}$$

An appropriate interpretation of this definition is that the component for which the component reliability imposes the greatest gradient on the system reliability function is the most important. Consider some specific example cases.

For a series structure comprised of n independent components, the realization of expression is

$$I_R(j) = \frac{d}{dr_j} R_s = \frac{d}{dr_j} \prod_{i=1}^{n} r_i = \prod_{\substack{i=1 \\ i \neq j}}^{n} r_i$$

Thus, for a series system of independent components, the reliability importance of any component is equal to product of the reliabilities of the other components. Since component indices are usually arbitrary, assume the components have been numbered so that

$$r_1 \le r_2 \le r_3 \le \dots\dots\dots \le r_n$$

Under this indexing, we can see that

$$I_R(1) \ge I_R(2) \ge \dots\dots\dots\dots \ge I_R(n)$$

so that the weakest component has the greatest importance. For a series system, this seems intuitively reasonable.

For a parallel system comprised of independent components, the corresponding analysis is

$$I_R(j) = \frac{d}{dr_j} R_s = \frac{d}{dr_j} \coprod_{i=1}^{n} r_i = \frac{d}{dr_j}\left(1 - \prod_{i=1}^{n}(1-r_i)\right) = \prod_{\substack{i=1 \\ i \neq j}}^{n}(1-r_i)$$

Assuming the same ordering of the component indices as indicated above, the most important component in a parallel system is the strongest. That is

$$I_R(1) \le I_R(2) \le \dots\dots\dots\dots \le I_R(n)$$

Intuitively, it seems reasonable that the strongest component is most important to reliability. The corresponding point that investments in component reliability improvement should begin with the most reliable component is less apparent but equally accurate.

Resolving questions of how to enhance system reliability has been one of the areas in which reliability analysts have contributed to system design efforts. Reliability importance measures provide a basis for evaluating the cost effectiveness of investments in component redesign or other improvement strategies.

3.5 RELIABILITY ALLOCATION

Another approach to enhancing system reliability is by introducing redundancy at selected component locations. That is, the system configuration is altered by replacing a single component with two or more copies of the component in parallel. The problem of selecting the components for which this is done is known as the reliability allocation problem. It is assumed that each copy of a component included in the system has a cost. The cost might actually represent the price of the component or may represent weight or any other consequence of allocating the component to the design. Then, the problem of designating the locations and magnitudes of component redundancy can be stated as an integer mathematical program. In fact, there are two plausible algebraic forms for the optimization problem. We may minimize system cost subject to a reliability constraint, or we may maximize system reliability subject to a budget constraint.

Regardless of the system configuration, we assume that we have an expression for the system reliability, R_s. Let m_i represent the number of copies of component i placed in parallel at the component i location in the system configuration and let c_i represent the unit cost for component i. Then, one integer programming representation of the system design problem is

$$\text{Minimize } \sum_{i=1}^{n} c_i m_i$$

subject to

$$R_s \geq R_{target}$$
$$m_i \geq 1 \quad \forall i$$
$$m_i \text{ integer}$$

Here, we minimize the cost to obtain a target system reliability level. The alternate problem is to

$$\text{Maximize } R_s$$

subject to

$$\sum_{i=1}^{n} c_i m_i \le C_{budget}$$

m_i integer

Naturally, for both of the optimization problems, the algebraic statement of the system reliability function has terms of the form

$$1-(1-r_i)^{m_i}$$

for the contributions of the reliabilities of the component positions. Equally clear should be the fact that it is the algebraic form of the system reliability function that determines how difficult it is to solve either of the optimization problems. A recent paper by Majety, Dawande, and Rajgopal [14] provides an efficient algorithm for solving the integer programs in general, and recent work by Rice, Cassady, and Wise [15] suggests that, because of the relatively small number of feasible solutions, enumeration strategies will often yield solutions efficiently.

A distinction should be made here between reliability allocation and reliability apportionment. Reliability apportionment is the process of assigning reliability targets to subsystems during system design. There are algorithms for doing this, the most popular of which is called the ARINC method, which is described by Lloyd and Lipow [16]. Usually, the algorithms are based upon a balance between the reliability importance of the subsystems and the cost of enhancing the reliability of existing subsystems designs. The key difficulty with the apportionment task is that the assignment process is driven by subjective criteria, in that enhancement cost is usually represented by an effort function. The origin of the effort function is often unclear. For this reason, the algorithm is not developed here. In contrast, the allocation models defined above treat the question directly.

3.6 CONCLUSION

The analyses considered to this point provide a means for relating system reliability to component reliability for many types of equipment designs. Several exceptions have been noted, and some of these will be addressed later in the text. For the system configurations that are based on binary component states and independent components, the models and analyses treated so far are sufficient to permit a reductionist approach to reliability analysis. That is, for these simplest of systems, reliability may be studied at the component level because the dependence of system reliability on component reliability is well defined. For very many system designs, the ability to focus independently on individual component reliability performance is essential to achieving the high levels of reliability we now enjoy.

3.7 EXERCISES

1. For the system of Problem 5 of Chapter 2, assume that $\underline{r}$ = {0.96, 0.96, 0.96, 0.90, 0.90, 0.94, 0.94). Compute the three types of reliability bounds as well as actual system reliability.
2. For the system of Problem 6 of Chapter 2, assume that $r_1 = r_2 = 0.96$, $r_3 = 0.95$, $r_4 = 0.98$, $r_5 = r_6 = 0.92$, $r_7 = 0.93$, and $r_8 = r_9 = 0.90$. Compute the three types of reliability bounds as well as actual system reliability. Also, compute the lower bound under the modular decomposition you constructed for the system.
3. Compute the reliability of a 3-out-of-4 system for which all components have a reliability of 0.85.
4. Compute the reliability of a three-component parallel system in which all components have a reliability of 0.75.
5. Suggest an alternative reliability importance measure to the one of Definition 3.1.
6. For the system of Problem 7 of Chapter 2, assume that $r_1 = 0.85$, $r_2 = 0.80$, $r_3 = 0.95$, $r_4 = 0.75$, and $r_5 = 0.90$. Compute the three types of reliability bounds as well as actual system reliability.

7. For the system of Problem 8 of Chapter 2, assume that $r_1 = r_2 = 0.90$, and $r_3 = r_4 = 0.80$. Compute the three types of reliability bounds as well as actual system reliability.

4

Reliability over Time

The definition of reliability given in Chapter 2 indicates that reliability is the probability of proper function over time. An implicit feature of this definition is the underlying assumption that, across a population of identical devices, survival over time (or life length) is dispersed in a manner that is well modeled by probability and hence by a probability distribution function. Thus, the extension of the measures of reliability to include time involves the specification of probability distributions that constitute reasonable models for the dispersion of life length. This is the subject of the present chapter.

4.1 RELIABILITY MEASURES

The logical extension of the models examined in the previous two chapters to include time is to indicate that component and system states may be represented as functions of time. That is, we map x_i to $x_i(t)$, and ϕ to $\phi(t)$. The interpretation of this extension is that the random variable implied in Definition 2.1 is the duration of proper system function. In similar manner, we adjust the definitions of the reliability functions as

$$R_s(t) = \Pr[\phi(t) = 1]$$

and

$$r_i(t) = \Pr[x_i(t) = 1]$$

As before, the complement of reliability is failure.

As indicated above, the introduction of the time variable carries the implicit assumption that the time at which a component or system state changes from proper function to failed, the failure time, is a random variable, and across a population of identical devices, the dispersion in the failure time can be represented using a probability distribution function. The failure time of a device is also called its life length. Denoting life length by T, the distribution function on T is represented by $F_T(t)$ where

$$F_T(t) = \Pr[T \leq t] = \text{the probability that device life length is less than or equal to } t$$

Given this basic definition, reliability in time is the probability that device life length exceeds t. Thus,

$$R(t) = \bar{F}_T(t) = \text{the probability that life length exceeds } t$$

These basic definitions raise two points concerning notation. First, note that it is reasonable to use either $R(t)$ or the survivor function $\bar{F}_T(t)$ to represent reliability. As these forms are truly synonymous, the survivor function form will be used here whenever reasonable. Second, note that the subscript denoting the component or system has been dropped. Much of the discussion to follow is general, in that it applies equally well to the system or to any component. As no specific component is being identified, the subscript is excluded unless needed for clarification. In addition, the comment at the end of Chapter 3 concerning the fact that our ability to reduce our focus from system to component in many designs suggests that we can often examine reliability independent of component identity.

Returning to the inclusion of time in our models, note that the distribution function on life length is the basis for four equivalent algebraic descriptors of longevity. These four descriptors are the distribution function, the survivor (reliability) function, the density function, and the hazard function. Reiterating the above definitions, we have

Distribution function: $F_T(t) = \Pr[T \le t]$ (4.1)

Reliability function: $\bar{F}_T(t) = \Pr[T > t]$ (4.2)

Density function: $f_T(t) = \frac{d}{dt} F_T(t)$ (4.3)

The hazard function, $z_T(t)$, is the instantaneous conditional probability of failure given survival to any time. That is, it is the instantaneous rate of failure for surviving devices. Algebraically,

$$\Pr[T \le t + \Delta t \mid T > t] = \frac{F_T(t+\Delta t) - F_T(t)}{\bar{F}_T(t)}$$

so we obtain the hazard function as the limit of this expression.

$$\begin{aligned} z_T(t) &= \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} \Pr[T \le t + \Delta t \mid T > t] \right\} \\ &= \lim_{\Delta t \to 0} \left\{ \frac{1}{\Delta t} \frac{F_T(t+\Delta t) - F_T(t)}{\bar{F}_T(t)} \right\} = \frac{f_T(t)}{\bar{F}_T(t)} \end{aligned} \quad (4.4)$$

As indicated by the algebraic form, the hazard function is the rate at which surviving units fail. For this reason, it is often called the "failure rate." However, because it can apply to other failure phenomena, the terminology "failure rate" can be misleading, so in this text, the function $z_T(t)$ is called the hazard function.

Knowledge of any one of the four reliability measures implies knowledge of all of them. They are all functionally related and actually comprise alternate statements of the model of life length dispersion. As an example, note that given $f_T(t)$, we obtain

$$F_T(t) = \int_0^t f_T(u)du$$

$$\bar{F}_T(t) = 1 - F_T(t)$$

and

$$z_T(t) = \frac{f_T(t)}{\bar{F}_T(t)}$$

As a further example, suppose we are given $z_T(t)$. Treating the definition of the hazard function

$$z_T(t) = \frac{f_T(t)}{\bar{F}_T(t)}$$

as a differential equation, we have

$$f_T(t) = -\frac{d}{dt}\bar{F}_T(t) = z_T(t)\bar{F}_T(t)$$

so

$$\frac{d}{dt}\bar{F}_T(t) + z_T(t)\bar{F}_T(t) = 0$$

The solution for this equation is

$$\bar{F}_T(t) = e^{-\int_0^t z_T(u)du}$$

or alternately,

$$\bar{F}_T(t) = e^{-Z_T(t)}$$

where

$$Z_T(t) = \int_0^t z_T(u)du \tag{4.5}$$

is called the cumulative hazard function. Later in this text, there are several topics in which the cumulative hazard is a useful part of the analysis.

As the four reliability measures are all faces of the same description of the failure behavior of a device, we could use any of them as a basis for distinguishing failure patterns. The hazard function is commonly used by reliability analysts to describe the failure behavior of a device. The use of the hazard function started with the concept that a population of devices displays a "bathtub-shaped" hazard over the lives of the members of the population. The "bathtub curve" is shown in Figure 4.1. The shape is intended to illustrate the view that aging in a device population proceeds through phases. Early in the lives of the devices, failures occur at a relatively high rate. This "infant-mortality period" is often attributed to the failure of members of the population that are "weak" as a result of material flaws, manufacturing defects, or other physical anomalies. Following the "early-life" or "infant-mortality" period, the device population proceeds through the "functional-life period" during which the hazard function is relatively low and reasonably stable. Finally, toward the end of the lives of the population members, survivors fail with an increasing rate as a consequence of "wear out."

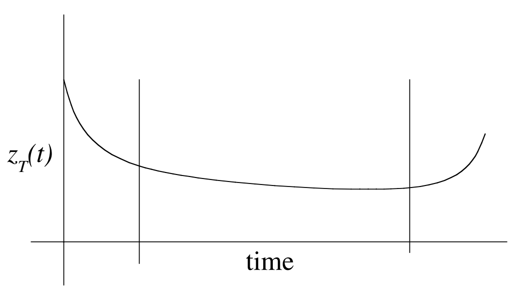


Figure 4.1 Example of a bathtub curve.

It is reasonable to observe that actuarial curves for human and other biological entities often display the bathtub shape, so that the analogy to human mortality is often informative. It is also interesting that early-life failure behavior has been observed so extensively that most durable goods manufacturers include some sort of run-in as part of their product testing activities. In addition, it has long been common for government and military procurement policies to mandate run-in efforts as a condition of sale for equipment suppliers.

The concept of the bathtub curve has been discussed and debated widely by reliability analysts. Some authors such as Wong and Lindstrom [17] argue that device populations are actually comprised of numerous subpopulations, each of which has a unique hazard behavior. According to Wong and Lindstrom, mixing the subpopulations results in a hazard curve that is "roller coaster" rather than bathtub shaped. The key point is that the hazard function is viewed as the most informative descriptor of device failure behavior. It is the measure that is usually used to select the distribution function to model life length.

In view of the importance of the hazard functions to descriptions of device failure patterns, an extensive classification of hazard function behavior has been devised. The simplest elements of that classification are presented below.

Defn. 4.1: A life distribution, $F_T(t)$, is said to be an *increasing failure rate* (IFR), distribution if

$$\frac{d}{dt} z_T(t) \geq 0 \ , \ \ 0 \leq t < \infty \tag{4.6}$$

An alternate condition for the IFR classification is that

$$\bar{F}_T(t+\tau \mid \tau) = \frac{\bar{F}_T(t+\tau)}{\bar{F}_T(\tau)}$$

be nonincreasing in τ for all $t \geq 0$. Note that in the above definition and the ones that follow, the terminology "failure rate" is used. This is done to show the correspondence to the abbreviations that were defined by Barlow, Marshall, and Proschan [18] at a time when the ambiguity in the terminology had not yet been recognized.

In any case, the above expression says that, if the conditional survival probability is a nonincreasing function of age, then the rate at which failures occur is increasing and the life distribution is IFR. By similar reasoning:

> *Defn. 4.2*: A life distribution, $F_T(t)$, is said to be a *decreasing failure rate* (DFR), distribution if

$$\frac{d}{dt} z_T(t) \leq 0 \ , \quad 0 \leq t < \infty \tag{4.7}$$

The alternate condition for the DFR classification is that $\bar{F}_T(t+\tau \mid \tau)$ be nondecreasing in τ for all $t \geq 0$. The third possible form for the hazard function is

> *Defn. 4.3*: A life distribution, $F_T(t)$, is said to be a *constant failure rate* (CFR), distribution if

$$\frac{d}{dt} z_T(t) = 0 \ , \quad 0 \leq t < \infty \tag{4.8}$$

The equivalent condition for a CFR classification is that

$$\bar{F}_T(t+\tau \mid \tau) = \bar{F}_T(t)$$

for all $t \geq 0$. This is an interesting special case that is examined further in the next section.

There are situations in which the conditions for designation as IFR or DFR are only partially met. For these cases, we have

> *Defn. 4.4*: A life distribution, $F_T(t)$, is said to be an *increasing failure rate on average* (IFRA), distribution if

$$z_T(t) > \frac{1}{t}\int_0^t z_T(u)du \ , \quad 0 \leq t < \infty \tag{4.9}$$

and is said to be a *decreasing failure rate on average* (DFRA), distribution if

$$z_T(t) < \frac{1}{t}\int_0^t z_T(u)du\ ,\ \ 0 \le t < \infty \tag{4.10}$$

While more detailed and extensive classifications of hazard functions have been defined, the ones enumerated above are sufficient for our study here. The classification of the hazard behavior serves to direct our choice of distribution to model device reliability.

4.2 LIFE DISTRIBUTIONS

In principle, any distribution function may be used to model equipment longevity. In practice, distribution functions having monotonic hazard functions seem most realistic, and within that class, there are a few that are generally thought to provide the most reasonable models of device reliability. The most common choices of life distribution models are described in the next few pages.

4.2.1 The Exponential Distribution

The most widely used distribution function for modeling reliability is the exponential distribution. It is such a popular model of device reliability because (1) it is algebraically simple and thus tractable, and (2) it is considered representative of the functional life interval of the device life cycle. Some firms try to manage components or devices by aging them through the early life period before putting them in service. The devices are expected to be obsolete before reaching the wear-out period, so an appropriate model of device reliability is one having constant hazard. This point of view is controversial. Nevertheless, the exponential model is widely used. The general statement of the exponential distribution is

$$F_T(t) = 1 - e^{-\lambda t} \tag{4.11}$$

The corresponding density function is

$$f_T(t) = \lambda e^{-\lambda t} \tag{4.12}$$

and the hazard function is constant over time. It is

$$z_T(t) = \frac{f_T(t)}{\bar{F}_T(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda \tag{4.13}$$

The exponential distribution is the only probability distribution that has a constant hazard function. In fact, if we observe that a device appears to have constant hazard, we may conclude that the exponential distribution is an appropriate life distribution model and the converse statement also applies. To see that constant hazard implies an exponential distribution, solve the differential equation implied by Equation 4.13:

$$\frac{d}{dt}F_T(t) = f_T(t) = \lambda \bar{F}_T(t)$$

and obtain Equation 4.11 as the result.

Constant hazard is both a desirable and an undesirable feature of the exponential model. The appeal of the result lies in its simplicity. Problems with the constant hazard model revolve around the associated "memoryless" property that it displays and the corresponding fact that the conditional survival probability is independent of age. That is

$$\Pr[T > t + \tau \mid T > \tau] = \bar{F}_T(t + \tau \mid \tau) = \frac{\bar{F}_T(t+\tau)}{\bar{F}_T(\tau)} = \frac{e^{-\lambda(t+\tau)}}{e^{-\lambda\tau}} = e^{-\lambda t} = \bar{F}_T(t)$$

The interpretation of this result is that a used device has the same reliability as a new one. Clearly, this is quite contrary to intuition and is unlikely to be true of most devices. The lack of memory feature of the exponential model is therefore a weakness in its representation of real equipment.

One final observation concerning the exponential model is the fact that the life distribution of a series system comprised of independent components each of which has an exponential life distribution is exponential. That is

$$R_s(t) = \prod_{i=1}^{n} \bar{F}_{T_i}(t) = \prod_{i=1}^{n} e^{-\lambda_i t} = e^{-\sum_{i=1}^{n}\lambda_i t} = e^{-(\sum_{i=1}^{n}\lambda_i)t}$$

Note further that the above expression confirms that fact that the system level hazard function for a series system of independent components is computed as the sum of the component hazards.

4.2.2 The Weibull Distribution

An alternate life distribution model that is also widely used is the Weibull distribution. The distribution is named for its developer, Waloodi Weibull, who was a highly talented Swedish scientist. He developed the Weibull distribution [19, 20] to describe the observed strengths of tensile test specimens. It has subsequently been found that the distribution provides a reasonable model for the life lengths of very many devices. The Weibull distribution function may be stated in several ways. The most general is

$$F_T(t) = 1 - e^{-\left(\frac{t-\delta}{\theta-\delta}\right)^{\beta}} \tag{4.14}$$

For this three-parameter form of the distribution function, the parameter δ is a minimum-life parameter that is often assumed to have a value of zero.

The interpretation of the parameter δ is that it is the time before which no failures occur. When expressed in this manner, it seems reasonable to set $\delta = 0$. On the other hand, if the "time variable" is actually cycles to failure or applied force in the case of a mechanical component, $\delta > 0$ may be an appropriate feature of the failure model. For example, a tensile specimen made of steel will not fail when subjected to forces of 10 to 20 kg/cm^2. To model the dispersion in failure strength of such specimens, a minimum applied force of perhaps $\delta = 100$ kg/cm^2 might be appropriate. When the minimum

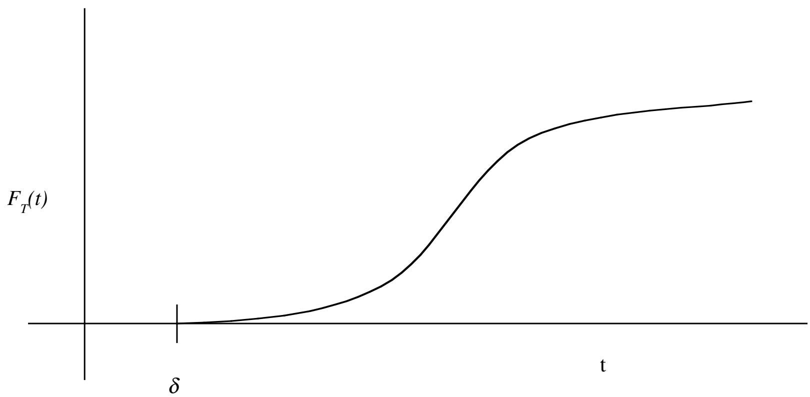


Figure 4.2 Three-parameter Weibull life distribution.

life parameter is nonzero, the distribution function appears as is shown in Figure 4.2.

Because it is simply a coordinate location parameter, there is no loss in generality in assuming it has value zero. For the balance of the discussion here, we take $\delta = 0$. Then, the form of the two-parameter Weibull distribution function is

$$F_T(t) = 1 - e^{-\left(\frac{t}{\theta}\right)^{\beta}} \tag{4.15}$$

The parameter θ is a scale parameter in that it determines the range of dispersion of the distribution. It is also called the "characteristic life" parameter because the value of the distribution at $t = \theta$ is independent of the value of the second parameter, β. That is

$$F_T(t = \theta) = 1 - e^{-1} = 0.632$$

so θ is a characteristic feature of any realization of the distribution.

The parameter β is the shape parameter. It determines the relative shape of the distribution, and it also determines the behavior of the hazard function. The general form of the hazard function is

$$z_T(t) = \frac{\beta t^{\beta-1}}{\theta^{\beta}} \tag{4.16}$$

Figure 4.3 provides an illustration of the Weibull hazard function. Observe that by setting $\beta = 1$, we obtain the exponential distribution as a special case and thus have constant hazard. Note further that the hazard is monotonically increasing when $\beta > 1$ and monotonically decreasing when $\beta < 1$.

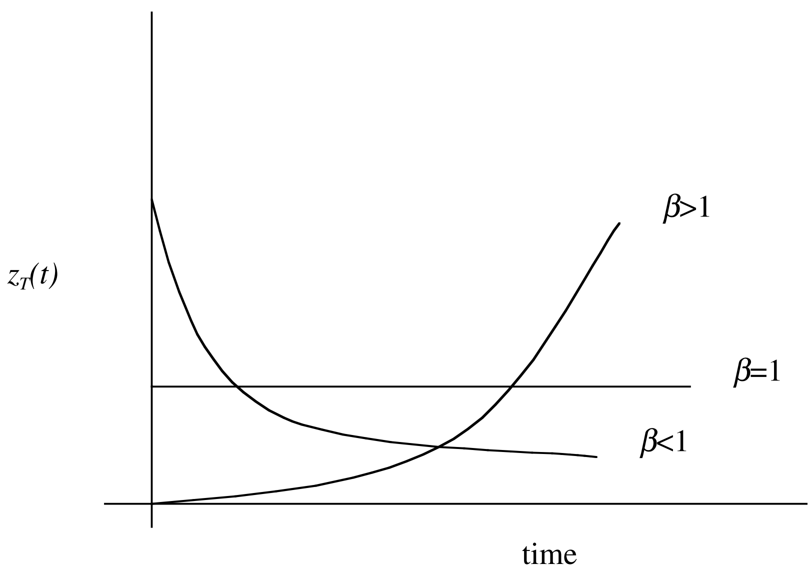


Figure 4.3 Weibull hazard functions.

The Weibull distribution is very widely used in reliability modeling. It has the advantages of flexibility in modeling various types of hazard behavior and of algebraic tractability. In addition, as with any two parameter distribution, it can be made to fit many actual situations reasonably well.

There are two further reasons that the Weibull is so widely used. One is that when Weibull [19] first developed the distribution form, it was to represent the failure behavior of tensile specimens, and the other is that the Weibull is one possible realization of the extreme value distribution. Considering that a device is subject to failure due to any of several causes, the first failure mechanism to be activated (smallest time to occurrence) determines device failure. Thus, failure times are smallest extremes of a set and might reasonably be modeled using a distribution from the class of extreme value distributions.

Another perspective on the extreme value concept is that one might consider that there are some number, say k, of possible sites (links in a chain, units of material, reactive species, etc.) of failure, and that each cause is actuated in time according to an identical and independent distribution. The time of device failure is the minimum of the cause actuation times. This also implies an extreme value. The extreme value argument may be stated as

$$F_T(t) = F_{T_j}(t) \prod_{\substack{i=1 \\ i \neq j}}^{k} \bar{F}_{T_i}(t) = F_{T_j}(t)\left(\bar{F}_{T_j}(t)\right)^{k-1}$$

where the subscripts on the time variable indicate a correspondence to the pertinent "units." Gumbel [21] presents results developed by Fisher and Tippet [22], which show that, in the limit as k goes to infinity, the life distribution can only have one of three forms, and the forms correspond to cases of negative, unconstrained in sign, and positive random variables. The form for nonnegative random variables has the Weibull as a representative case. In addition, the variable $Y = \ln T$, where T has a Weibull distribution, has an extreme distribution of the form

$$F_Y(y) = 1 - e^{-e^{\beta(y - \ln\theta)}} \tag{4.17}$$

Thus, the Weibull has a fairly plausible physical interpretation.

There are two special cases of the Weibull that are used fairly extensively. One is when the value of the shape parameter β is 3.26. At this value, the Weibull is nearly "bell shaped," so it appears like and can be used as a substitute for the Normal distribution. The other special case is when the value of the shape parameter β is 2. In this case, the distribution is usually called the Raleigh distribution. It is special because the hazard function is linear:

$$z_T(t) = \frac{2t}{\theta^2}$$

and many people find this linearity appealing.

4.2.3 The Normal Distribution

Another popular model of device life length is provided by the Normal distribution. Because of its algebraic intractability, the Normal is usually expressed in terms of its density function:

$$f_T(t) = \frac{e^{-(t-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \tag{4.18}$$

The corresponding distribution function cannot be expressed algebraically in closed form. Hence, the same statement applies to the reliability and hazard function. The distribution function is widely tabulated in its standard form. A table of standard normal probabilities is included in Appendix A. It is also reasonably straightforward to compute Normal variates and cumulative probabilities numerically. The expressions for these calculations are provided in Appendix A.

The parameters of the Normal distribution correspond to the distribution moments. That is, $\mu = \mathrm{E}[t]$ is the mean of the distribution, and σ^2 is the variance. The Normal distribution displays monotonically increasing hazard. We can prove that fact as follows. Start with the definition of the hazard function as given in Equation 4.4 and take the derivative of the hazard function to obtain

$$\frac{d}{dt} z_T(t) = \frac{f_T'(t)\left\{\bar{F}_T(t)\right\} - f_T(t)\left\{-f_T(t)\right\}}{\left\{\bar{F}_T(t)\right\}^2} = \frac{\bar{F}_T(t) f_T'(t) + f_T^2(t)}{\bar{F}_T^2(t)}$$

Next, take the indicated derivative of the density function

$$f'_T(t) = \frac{d}{dt}\frac{e^{-(t-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = -\frac{2(t-\mu)}{2\sigma^2}\frac{e^{-(t-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = -\frac{(t-\mu)}{\sigma^2} f_T(t)$$

and substitute it back into the derivative of the hazard function:

$$\frac{d}{dt} z_T(t) = \frac{-\frac{(t-\mu)}{\sigma^2} f_T(t)\bar{F}_T(t) + f_T^2(t)}{\bar{F}_T^2(t)}$$

Note that

$$\bar{F}_T^2(t) > 0 \quad \text{so} \quad \frac{d}{dt} z_T(t) \geq 0 \quad \text{if} \quad -\frac{(t-\mu)}{\sigma^2} f_T(t)\bar{F}_T(t) + f_T^2(t) \geq 0$$

When

$$t < \mu$$

$$-\frac{(t-\mu)}{\sigma^2} > 0$$

$$f_T(t) \geq 0$$

$$\bar{F}_T(t) > 0$$

and

$$f_T^2(t) \geq 0$$

so

$$\frac{d}{dt} z_T(t) \geq 0$$

When

$$t \geq \mu$$

$$-\frac{(t-\mu)}{\sigma^2} f_T(t)\bar{F}_T(t) + f_T^2(t) = \left(-\frac{(t-\mu)}{\sigma^2}\bar{F}_T(t) + f_T(t)\right) f_T(t)$$

and

$$f_T(t) \geq 0$$

So the question is whether or not

$$-\frac{(t-\mu)}{\sigma^2}\bar{F}_T(t) + f_T(t) \geq 0$$

To decide this, we recall that

$$\bar{F}_T(t) = \int_t^\infty f_T(x)dx$$

Necessarily, for any fixed value t,

$$t\int_t^\infty f_T(x)dx < \int_t^\infty x f_T(x)dx$$

Therefore,

$$\frac{(t-\mu)}{\sigma^2}\bar{F}_T(t) = \frac{(t-\mu)}{\sigma^2}\int_t^\infty f_T(x)dx < \int_t^\infty \frac{(x-\mu)}{\sigma^2} f_T(x)dx$$

$$= \int_t^\infty \frac{(x-\mu)}{\sigma^2}\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}dx = -f_T(x)\Big|_t^\infty = f_T(x)$$

and hence

$$\frac{(t-\mu)}{\sigma^2}\bar{F}_T(t) < f_T(x)$$

This inequality implies that

$$-\frac{(t-\mu)}{\sigma^2}\bar{F}_T(t) + f_T(t) \geq 0$$

so the derivative of the Normal hazard function is nonnegative and the hazard function is increasing.

As is discussed in Chapter 5, the Normal distribution is often considered a very appropriate model for the reliability of structural components. On the other hand, many reliability analysts have resisted the use of the Normal distribution as a life length model for time indexed ages because of the fact that the Normal is defined over the entire real line, negative as well as positive.

4.2.4 The Lognormal Distribution

A suggested response to the criticisms of the Normal distribution is the use of the lognormal. In this case, ln t is taken to be normally distributed. As a consequence, the density function for the life distribution model becomes

$$f_T(t) = \frac{e^{-(\ln t - \mu)^2/2\sigma^2}}{t\sqrt{2\pi\sigma^2}} \tag{4.19}$$

Note the appearance of the time variable in the denominator of this function.

It also appears that the lognormal is an appealing distribution independent of the issues related to positive random variables. The life lengths of quite a few microelectronic components have been found to be well modeled by the lognormal distribution.

The lognormal distribution suffers from a comparable level of algebraic intractability to the Normal. The model is quite useful, and it displays the unique feature that for appropriately selected parameter values, the hazard function increases and then decreases.

4.2.5 The Gamma Distribution

Another distribution that is widely used in reliability modeling is the Gamma distribution. The representation of the density function for the Gamma is

$$f_T(t) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t} \tag{4.20}$$

The distribution function can be stated in closed form only if the shape parameter, β, is an integer. In this case, the distribution function is

$$F_T(t) = \sum_{k=\beta}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} = 1 - \sum_{k=0}^{\beta-1} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \tag{4.21}$$

Note that when the shape parameter is an integer, the Gamma distribution is usually called the Erlang (or Erlang-β) distribution.

As in the case of the Weibull, the Gamma distribution model displays increasing hazard when $\beta > 1$, decreasing hazard when $\beta < 1$, and constant hazard when $\beta = 1$. Thus, setting $\beta = 1$ collapses the Gamma distribution to the exponential.

The Gamma has the disadvantage of being rather difficult algebraically, but it has the advantage that it arises naturally as the convolution of identical exponential distributions. It therefore has considerable practical appeal. Strategies for the numerical evaluation of Gamma functions and Gamma distributions are provided in Appendix A.

4.2.6 Other Distributions

A wide variety of other probability distributions are used for reliability modeling but most with relatively low frequencies. One of those other distributions, the Birnbaum-Saunders distribution [23, 24], is discussed briefly in Chapter 5. Another is the Makeham [25] distribution, for which the distribution function is

$$F_T(t) = 1 - e^{-[\alpha t + \frac{\beta}{\gamma}(e^{\gamma t} - 1)]} \tag{4.22}$$

The hazard function for this distribution is

$$z_T(t) = \alpha + \beta e^{\gamma t} \tag{4.23}$$

so there are three parameters that can be selected to provide whatever type of model is desired. More importantly, the parameters can be selected so that the function matches failure data quite well. For this reason, the Makeham distribution [25] is widely used in actuarial studies. Note that it also corresponds to an extreme value type of distribution.

Still another model that is based on actuarial data analysis is the Gompertz distribution [25], for which the hazard function is

$$\lambda\beta^t \tag{4.24}$$

so the distribution function is

$$F_T(t) = 1 - e^{-\left[\frac{\lambda(\beta^t - 1)}{\ln\beta}\right]} \tag{4.25}$$

While this form is rather intricate, the rationale for its construction is that the reciprocal of the hazard function should be decreasing. That is,

$$\frac{d}{dt}\left(\frac{1}{z_T(t)}\right) < 0$$

which implies increasing hazard. For those working with actuarial data, this seems a reasonable way to treat such behavior, and the distribution has been adopted by some reliability specialists for the same reason.

One final model that is particularly worth mentioning is the one suggested by Hjorth [26]. He calls it the IDB distribution because, depending on the choice of parameter values, it can have increasing, decreasing, or bathtub-shaped hazard function. The general statement of the IDB distribution is

$$F_T(t) = 1 - \frac{e^{-\delta t^2/2}}{(1+\beta t)^{\theta/\beta}} \tag{4.26}$$

and the corresponding hazard function is

$$z_T(t) = \delta t + \frac{\theta}{1+\beta t} \tag{4.27}$$

Clearly, with three parameters, many behaviors can be modeled. In particular, setting $\theta = 0$ yields the Rayleigh type of Weibull distribution. Similarly, setting $\delta = \beta = 0$, yields the exponential distribution. Observe that $\delta = 0$ implies decreasing hazard, while $\delta > \theta\beta$ implies increasing hazard, and $0 < \delta < \theta\beta$ yields a bathtub-shaped hazard function. If only for its flexibility, the model is worthy of consideration.

In summary, we might note that there are very many types of equipment for which life distributions provide a meaningful model of life duration. The possible choices are comparably wide. The distributions described above are the principle but not the only distributions used to model life length. Each has advantages, and each has shortcomings. The key is to select one that is appropriate for its application.

4.3 SYSTEM LEVEL MODELS

The discussion of equivalent structures in Chapter 2 assures that system reliability may be expressed in terms of component reliability. In addition, the discussion in Chapter 3 of the use of minimum paths and minimum cuts to construct bounds on system reliability further supports the focus on component-level reliability measures. Notwithstanding this fact, system-level (or subsystem) analyses are often meaningful or informative. For example, consider a series system of independent components. For this system, the reliability function is

$$\bar{F}_T(t) = \prod_{i=1}^{n} \bar{F}_{T_i}(t) = \prod_{i=1}^{n} e^{-Z_{T_i}(t)} = e^{-\sum_{i=1}^{n} Z_{T_i}(t)} \tag{4.28}$$

Clearly, this form implies that for a system of independent components arranged in series, the system-level hazard function is the sum of the component hazard functions. Thus, system hazard may be managed by reducing component hazard function levels.

Another point of interest is the question of whether or not component hazard function behavior is preserved in the formation of systems. For example, is a system comprised of independent IFR components IFR? The answer is no. Such a system is IFRA but not IFR. The most well-known example of the accumulation of IFR behavior is the parallel arrangement of three independent components, all of which are CFR (which may be viewed as simultaneously IFR and DFR). For this system, the reliability is

$$\bar{F}_T(t) = \coprod_{i=1}^{3} \bar{F}_{T_i}(t) = 1 - \prod_{i=1}^{3}\left(1 - \bar{F}_{T_i}(t)\right) = 1 - \prod_{i=1}^{3}\left(1 - e^{-\lambda_i t}\right)$$

$$= e^{-\lambda_1 t} + e^{-\lambda_2 t} + e^{-\lambda_3 t} - e^{-(\lambda_1+\lambda_2)t} - e^{-(\lambda_1+\lambda_3)t} - e^{-(\lambda_2+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}$$

so the corresponding density function on life length is

$$f_T(t) = \lambda_1 e^{-\lambda_1 t} + \lambda_2 e^{-\lambda_2 t} + \lambda_3 e^{-\lambda_3 t} - (\lambda_1 + \lambda_2)e^{-(\lambda_1+\lambda_2)t}$$
$$- (\lambda_1 + \lambda_3)e^{-(\lambda_1+\lambda_3)t}$$
$$- (\lambda_2 + \lambda_3)e^{-(\lambda_2+\lambda_3)t} + (\lambda_1 + \lambda_2 + \lambda_3)e^{-(\lambda_1+\lambda_2+\lambda_3)}$$

and the hazard function is the ratio of these two functions. A plot of the hazard function for the case in which the parameters are normalized to sum to one ($\lambda_1 = 0.6$, $\lambda_2 = 0.3$, $\lambda_3 = 0.1$) is shown in Figure 4.4.

Naturally, the specific behavior of the hazard function depends upon the values of the parameters of the life distributions of the components and upon the specific system structure. Nevertheless, we may conclude that IFR behavior is not preserved when IFR components are combined to form a system. The same is true for DFR components.

In view of the above results concerning the aggregation of IFR components, it is clear that, for any system, the behavior of the system-level hazard function and the system-level reliability function should be examined carefully. The approach most likely to lead to a successful investigation of the hazard function at the system level is to form the system reliability function and to differentiate it, either algebraically or numerically, and to then evaluate the expression

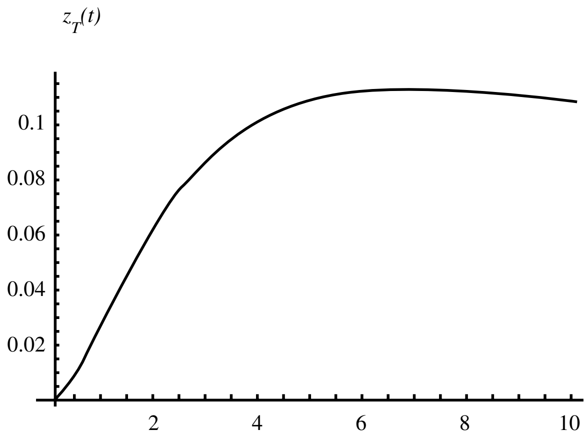


Figure 4.4 Hazard function for a three (exponential) component parallel system.

$$z_T(t) = \frac{-\frac{d}{dt}\bar{F}_T(t)}{\bar{F}_T(t)} \tag{4.29}$$

again either algebraically or numerically.

As an example, consider the system shown in Figure 4.5. This simple system has the following reliability function:

$$\bar{F}_T(t) = e^{-Z_{T_1}(t)}\left(e^{-Z_{T_2}(t)} + e^{-Z_{T_3}(t)} - e^{-Z_{T_2}(t)-Z_{T_3}(t)}\right)$$
$$= e^{-Z_{T_1}(t)-Z_{T_2}(t)} + e^{-Z_{T_1}(t)-Z_{T_3}(t)} - e^{-Z_{T_1}-Z_{T_2}(t)-Z_{T_3}(t)}$$

The density function corresponding to this reliability function is

$$f_T(t) = (z_{T_1}(t) + z_{T_2}(t))e^{-Z_{T_1}(t)-Z_{T_2}(t)} + (z_{T_1}(t) + z_{T_3}(t))e^{-Z_{T_1}(t)-Z_{T_3}(t)}$$
$$-(z_{T_1}(t) + z_{T_2}(t) + z_{T_3}(t))e^{-Z_{T_1}-Z_{T_2}(t)-Z_{T_3}(t)}$$

and the hazard function is the ratio of those two expressions. If all three components have Weibull life distributions with, $\theta_1 = 10$, $\beta_1 = 1.25$, $\theta_2 = \theta_3 = 8$, and $\beta_2 = \beta_3 = 2.25$, the system level hazard function is shown in Figure 4.6.

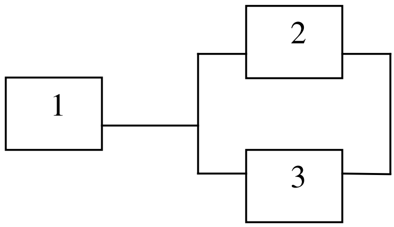


Figure 4.5 Example system.

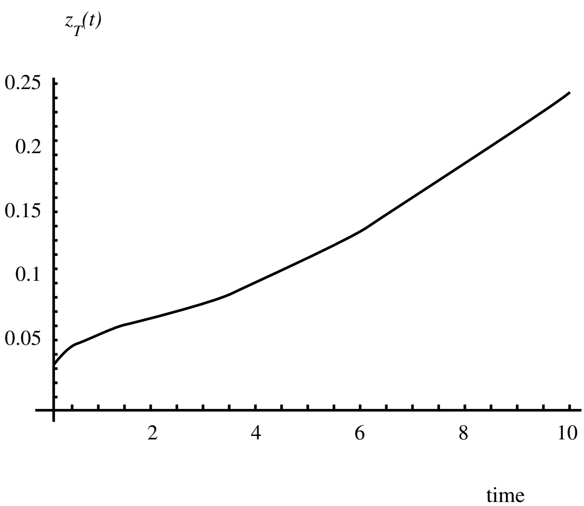


Figure 4.6 System-level hazard function.

Finally, before leaving the discussion of life distributions, let us consider the standby redundant component configuration. As mentioned in Chapter 2, this is a system structure that does not fall within the set of structural forms enumerated and is one that is used frequently. The system configuration has several possible realizations. Assuming that the system structure is represented by Figure 4.7, there are two components in parallel, but only one is performing the system task at a time. Component 1 operates until it fails, at which

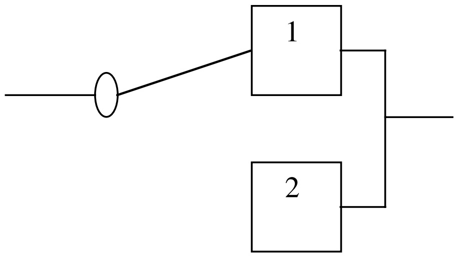


Figure 4.7 Standby redundant system.

time the switch instantaneously activates component 2. Then component 2 operates until it fails, at which time the system is failed. We may consider that the switch is perfect and always makes the change in components, or we may allow for the possibility that with probability, say p, the switch functions successfully, and with probability $(1 - p)$ the switch fails and the system is therefore failed. We may also consider either that component 2 is "warm" or "cold" while waiting to be activated. If we say that it is "warm," we imply that it is aging in some manner, while the statement that it is "cold" implies that component 2 is new when the switch moves to initiate its operation.

Now, the key quantity to determine is the distribution on the length of time the system operates before failure. That distribution is

$$F_{T_S}(t) = (1 - p)F_{T_1}(t) + p\int_0^t f_{T_1}(u)F_{T_2}(t-u)du \qquad (4.30)$$

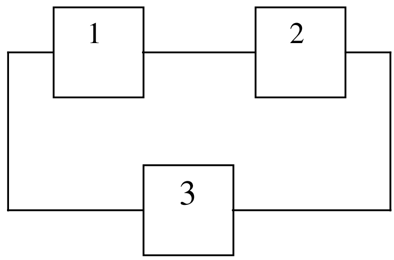
The rationale for this construction is that the system survives as long as the first component, if the switch fails, and as long as the sum of the failure times of the two components, if the switch functions properly. As a special case of this model, setting $p = 1$ represents the case of the perfect switch. One may also observe that, for this model, if $F_{T_1}(t)$ and $F_{T_2}(t)$ are both IFR, then the system life distribution is also IFR.

To conclude this chapter, it is appropriate to note that there are many other models that have been defined for particular types of equipment or specific types of operating profiles. The models discussed in this chapter represent the majority but not all of the useful models.

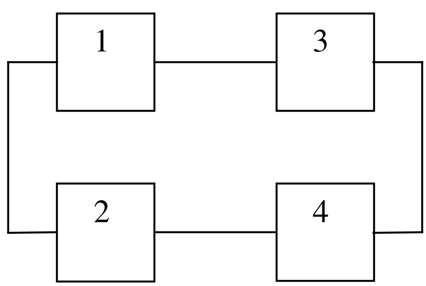
4.4 EXERCISES

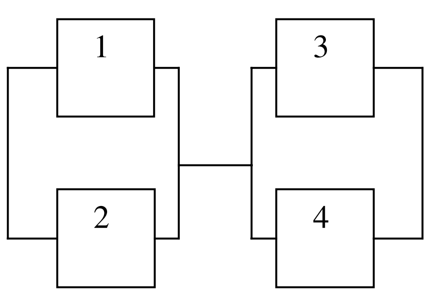
1. Plot the hazard function for a Gamma distribution having $\beta = 1.5$ and $\lambda = 0.005$.
2. Determine the point at which the hazard function for the lognormal distribution changes from increasing to decreasing.
3. Prove that the sum of Gamma-distributed random variables having the same scale parameter is Gamma distributed.
4. Prove that if a life distribution is IFR it is also IFRA, and that the converse is not true.
5. Construct an algebraic expression for the mean life of a parallel system of two independent components, each of which has an exponential life distribution.
6. Construct an algebraic expression for the reliability function and the system hazard function for a 2-out-of-3 system comprised of identical components each having an exponential life distribution. Plot the hazard function for the case in which $\lambda = 0.05$.
7. Compute the value of the reliability function at t = 500 hrs., t = 1000 hrs., and t = 12000 hrs. for a component population having Weibull life distribution with $\beta = 1.50$ and $\theta = 20000$. Also calculate the mean life length.
8. Repeat Problem 7 for a component population having exponential life distribution with $\lambda = 0.001$.
9. In general, for a distribution function $F_T(t)$, t_γ is the γ^{th} quantile of the distribution if $F_T(t_\gamma) = \gamma$. Determine the 0.90 quantile and the 0.99 quantile for the distributions in Problems 7 and 8.

10. Construct the algebraic expression for the system level hazard functions for the structure:



11. Construct algebraic expressions for the reliability and hazard functions for the following two systems. Compare several of the values of these functions for the case in which all components have Weibull life distributions and in both structures $\beta_1 = \beta_2 = 2.5$, $\theta_1 = \theta_2 = 100$, $\beta_3 = \beta_4 = 1.5$ and $\theta_3 = \theta_4 = 200$.





12. Analyze the standby redundant system for two identical exponential components having $\lambda = 0.001$ for the cases of $p = 0.50$, 0.80, and 1.0. Plot the resulting distribution function and its hazard function.
13. Consider a population of devices having life distribution $F_T(t)$. For items from this population that attain an age of, say, τ hours, denote by $U(\tau)$ the random variable representing their additional life lengths. Let $G_U(u(\tau))$ be the distribution function on $U(\tau)$. $G_U(u(\tau))$ is referred to as the residual life distribution and $\bar{G}_U(u(\tau))$ is its corresponding survivor function. Show that $F_T(t)$ is IFR if and only if $\bar{G}_U(u(\tau))$ is decreasing in τ for all u.
14. Construct the residual life distribution for exponential and Weibull life distributions.
15. Suppose a population of devices has a Weibull life distribution with $\beta = 1.6$ and $\theta = 25$. What is the mean of the residual life distribution for copies of the device that survive 15 hours?

5

Failure Processes

The fourth constituent of the definition of the reliability of equipment is the environment for which it has been designed. In saying "environment," we are really referring to the forces that are imposed on a device and the processes that lead to its deterioration and failure. We start with the question: Why does equipment fail? Many people agree that equipment fails because specific components fail and components fail because the operation of a system implies the application of forces (energy) upon the system and its components. Sudden and/or excessive forces precipitate immediate failure, while usual forces induce and sustain the progress of various types of deterioration processes under which the component eventually can no longer resist the usual force levels.

While the preceding explanation seems reasonable, there has never really been any consensus on the causes of failure and how they should be modeled. This situation persists today. A substantial portion of the scientific and engineering effort that has constituted the evolution of the reliability discipline has been focused on the study and modeling of component degradation processes. Numerous theories and models have been proposed, and very many heated debates have centered around the various points of view concerning the nature of failure processes.

In the pages that follow, I will enumerate several of the principal models and identify some of the points of contention.

I will try to suggest reasonable choices among the models and will finally offer my view on how failure models should be defined. A key feature of the models that have been proposed is that they are usually defined separately for mechanical systems, as opposed to electrical/electronic systems. That is, models of failure have traditionally been developed from either a mechanical or an electrical (electronic) perspective. Reliability of mechanical equipment has often been viewed as a problem in structural integrity as influenced by applied loads and inherent strength. In contrast, the reliability of electrical devices has usually been viewed as dependent upon material stability despite exposure to hostile chemical reactions such as oxidation. It is only recently that some analysts have suggested that both types of reliability result from common classes of phenomena. It is my view that this distinction is unnecessary, because the failure processes for mechanical and for electrical devices are essentially the same. I believe that they correspond to the progress and completion of physico-chemical reactions that occur in the materials that comprise devices and that are driven by the transfer of energy within those materials.

5.1 MECHANICAL FAILURE MODELS

5.1.1 Stress-Strength Interference

An early and still popular representation of mechanical device reliability is the "stress-strength interference" model. Under this model, there is random dispersion in the stress, Y, that results from applied loads. The dispersion in the stress realized can be modeled by a distribution function, say $H_Y(y)$. Similarly, there is also random dispersion in inherent device strength, X, and this can be modeled by $G_X(x)$. Then, device reliability corresponds to the event that strength exceeds stress. That is

$$\begin{aligned}\bar{F} = \Pr[X > Y] &= \int_{-\infty}^{\infty}\int_{y}^{\infty} h_Y(y)g_X(x)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} h_Y(y)g_X(x)dydx\end{aligned} \tag{5.1}$$

for which, an equivalent statement is

$$\begin{aligned}\bar{F} &= \Pr[X > Y] = \int_{-\infty}^{\infty} h_Y(y)\big(1 - G_X(y)\big)dy \\ &= \int_{-\infty}^{\infty} H_Y(x)g_X(x)dx\end{aligned} \tag{5.2}$$

The failure probability is the complement of the reliability. This time-independent model has been studied extensively. Kapur and Lamberson [27] provide solutions for several different choices of G and H.

Since the Expression 5.2 is independent of time, the definition of a time-based reliability model centers on selecting the distributions G and H and on representing the time evolution of those distributions. One popular model of this type is based on the assumption that stress is Normal in distribution, with constant mean and variance, while strength is also Normal, but with declining mean and increasing variance. The idea is that the aging of a device results in a gradual decrease in mean strength and a gradual increase in the inconsistency (vaiability) in strength. In this case

$$\begin{aligned}\bar{F}_T(t) &= \Pr[X(t) > Y] = \int_{-\infty}^{\infty} \phi\left(\frac{y-\mu_y}{\sigma_y}\right)\left(1 - \Phi\left(\frac{y-\mu_x(t)}{\sigma_x(t)}\right)\right)dy \\ &= \int_{-\infty}^{\infty} \Phi\left(\frac{x-\mu_y}{\sigma_y}\right)\phi\left(\frac{x-\mu_x(t)}{\sigma_x(t)}\right)dx\end{aligned} \tag{5.3}$$

where μ_y and σ_y are the constant parameters of the Normal stress distribution, $\mu_x(t)$ and $\sigma_x(t)$ are the time-dependent parameters of the strength distribution, ϕ denotes the standard Normal density, and Φ represents the cumulative distribution for the standard Normal.

Clearly, the time evolution of the parameters of the strength distribution may be assigned any plausible form. For example, if the mean declines linearly in time and the standard deviation increases linearly in time, possible functions are

$$\mu_X(t) = \mu_X(0) - \alpha t \text{ and } \sigma_X(t) = \sigma_X(0) + \beta t$$

Consider an example realization of this model. Suppose the distribution on imposed stress has μ_y = 100 kg/cm^2 and σ_y = 2 kg/cm^2, and the distribution on device strength has $\mu_x(0)$ = 102.5 kg/cm^2 and $\sigma_x(0)$ = 1 kg/cm^2. Then the basic concept of the stress-strength interference model is well illustrated in Figure 5.1. Note that the distribution on stress lies below that on strength, so that in general, the strength values will probably exceed the stress values.

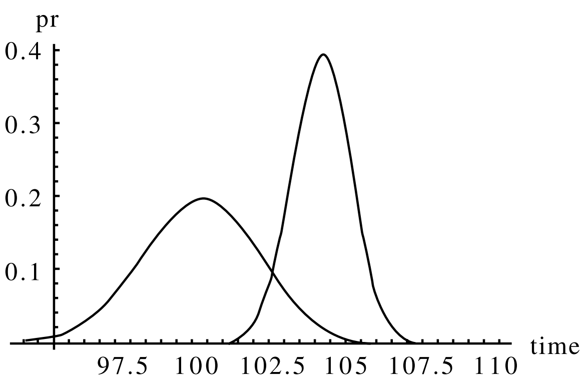


Figure 5.1 Basic stress-strength interference model.

Next, suppose that the "decay" parameters are α = 0.005 and β = 0.002. With these values, the gradual deterioration of the device strength is represented by a gradual change in the center and the width of the strength distribution. This is illustrated in Figure 5.2. At each point in time, the corresponding "slice" of the distribution corresponds to the strength distribution at that time. Consider the interference at time values of t = 250 and 500. The corresponding plots are shown in Figure 5.3. Clearly, as the mean of the strength distribution declines, the probability that strength exceeds stress also decreases. Since both the strength and the stress distributions have been assumed to be Normal, the values of realizations of the distribution and for the reliability must be computed

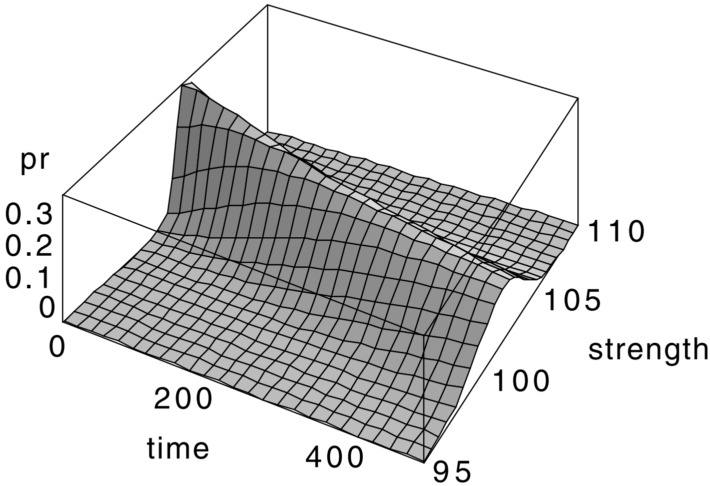


Figure 5.2 Time evolution of the strength distribution.

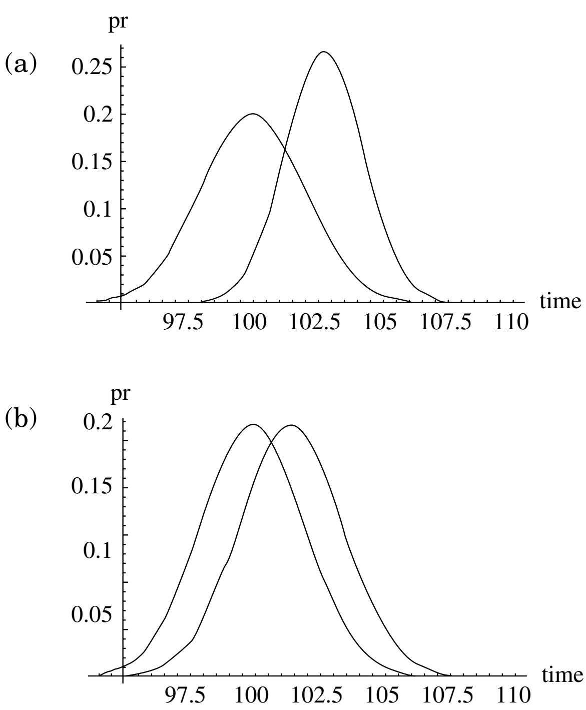


Figure 5.3 (a) Interference at t = 250. (b) Interference at t = 500.

numerically. Table 5.1 lists values of the reliability function of Equation 5.3 for the numerical example defined above.

A plot of the reliability function over the same range is shown in Figure 5.4. Once the numerical analysis of the model has been performed, one can fit a distribution to the reliability function as desired. For example, the function represented in Figure 5.4 (Table 5.1) is well represented by a Weibull distribution with parameters β = 1.083 and θ = 1204.

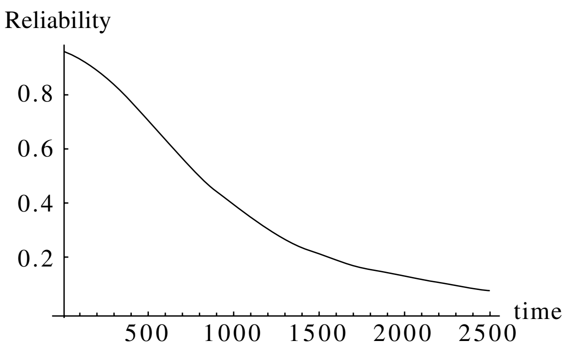


Figure 5.4 Example reliability function.

5.1.2 Shock and Cumulative Damage

An alternate and more widely used set of models are the shock and the cumulative damage models. An appealing description of the simplest shock model is offered by Gertsbakh and Kordonskiy [28], who was one of the pioneers in the development of reliability theory in Europe. Gertsbakh and Kordonskiy suggest that we consider a sequence of equipment actuations (or events), each of which imposes a stress on a component of interest. As long as the stress is below a threshold, the component does not fail, and the first time the stress exceeds the threshold, the component fails. As an example, he considers that each time an airplane lands, the landing imposes a gravitational force on the communications radio and that variations in weather conditions and pilot skill imply substantial variation in the loads experienced. When a sufficiently severe

Table 5.1 Example Reliability Function Values

Time	100	200	500	750	1000	1500	2000	2500	3000	4000	5000
Reliability	0.933	0.890	0.702	0.531	0.391	0.217	0.132	0.088	0.064	0.039	0.028

load occurs, the radio is damaged and fails. The life of the radio may thus be measured in numbers of landings and if γ is the probability that failure occurs on any landing, the distribution on life length is geometric, so

$$F_K(k) = \gamma \sum_{n=1}^{k} (1-\gamma)^{n-1} = 1-(1-\gamma)^k \tag{5.4}$$

gives the cumulative probability of failure on or before the k^{th} landing. One may then argue that for any individual landing, γ is quite small, so

$$(1-\gamma)^k \approx e^{-k\gamma}$$

or alternately, one may observe that

$$E[K] = 1/\gamma$$

so that the approximation

$$\bar{F}_K(k) \approx e^{-k/E[K]}$$

is reasonable. In either case, one obtains a life distribution model in terms of the number of cycles or actuations to failure that is exponential.

$$F_K(k) \simeq 1 - e^{-\gamma k} \tag{5.5}$$

Naturally, if we assume an average time between landings (or actuations), the model is easily converted to one expressed in terms of time. Of course, the designation and interpretation of a "time" scale is always somewhat arbitrary and can be defined according to the application.

Gertsbakh's example is informative in two ways. One is that it illustrates the fact that reliability models have usually been based on an engineering analysis of observed physical behavior. The second is that the example helps to emphasize the evolution that has occurred in modeling failure as engineers have made the models more and more representative of actual operating behavior. Starting from the above model, the transition across the cumulative damage models to the diffusion models show this evolution.

The early development of cumulative damage models was performed by Birnbaum and Saunders [24] in their study of fatigue in aircraft and by Karlin [29]. The elementary form of the model starts with the assumption that a device is subjected to "shocks" that occur randomly in time. Each shock imparts a random quantity, say X_i, of damage to the device, which fails when a capacity or endurance threshold is exceeded. The most common realization of this model includes the assumption that the shocks occur according to a Poisson process with intensity λ, and the amounts of damage per shock are independently and identically distributed according to some arbitrarily selected common distribution, say G. If $\bar{F}_T(t)$ represents the reliability over time, and K is the number of shocks that occur over the interval [0, t], the reliability function based on a threshold of L is

$$\bar{F}_T(t) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} G_X^{(k)}(L) \tag{5.6}$$

Note that the sum is taken over all possible numbers of shocks, and the notation $G_X^{(k)}(x)$ represents the k-fold convolution of $G_X(x)$ and thus the sum of k shock magnitudes, X. Thus, Equation 5.6 represents the probability that k shocks occur and their sum does not exceed the strength/damage threshold L — summed over all possible values of k. By convention, $G_X^{(0)}(x) = 1$ for all values of $x \geq 0$, and of course, $G_X^{(1)}(x) = G_X(x)$. Parenthetically, independent of the choice of distribution $G_X(x)$ for modeling shock magnitudes, Equation 5.6 will correspond to an IFRA distribution.

Consider an example. Assume that a device has an endurance threshold $L = \mu_x(0) = 102.5$ kg/cm², as in the case of the stress-strength interference model. Let $G_X(x)$ be a Normal distribution with $\mu_x = 50$ kg/cm² and $\sigma_x = 1$ kg/cm², and assume $\lambda = 0.004$ per hour. Solving Expression 5.6 as a function of time yields the reliability curve shown in Figure 5.5.

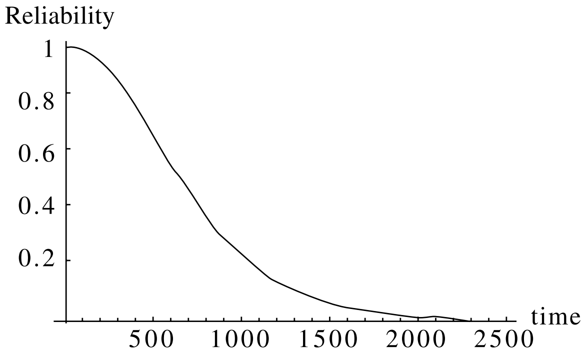


Figure 5.5 Reliability function for the cumulative damage model.

Note the conceptual duality of the cumulative damage model and the stress-strength interference model. In the stress-strength interference model, we treated stress as constant and strength as variable. In the basic cumulative damage model, the strength (damage threshold) is the constant quantity, and the stress (damage) is variable.

A special case of the cumulative damage model was suggested by Gertsbakh and Kordonskiy [28]. They considered that shocks occurred according to a Poisson process and were always of the same fixed magnitude. In this case, the model reduces to that of a birth process with a termination barrier, and the life distribution is a Gamma distribution. Also, Nachlas [30] generalized Gertsbakh's model by treating the damage rate as state dependent and the initial and failure states as random. The result is that the life distribution model becomes a generalized Gamma distribution. In the generalized model, the state dependent arrival rate is defined by the polynomial

$$\lambda_j(t) = j^n \alpha m t^{m-1} \tag{5.7}$$

for a component in state j. The state variable represents the degree of degradation of the device. For a failure threshold of state k and a random initial state, say i, the time to failure is described by

$$F_T(t\,|\,i,k) = 1 - \left(\prod_{l=i}^{k-1} l^n\right) \sum_{r=i}^{k-1} \left(\frac{1}{r^n} \prod_{\substack{l=i \\ l \neq r}}^{k-1} (l^n - r^n) \right) e^{-r^n \alpha t^m} \qquad (5.8)$$

This model is further enhanced by taking a Poisson mixture over the distance (k–i) between the initial and final state.

Subsequent to its initial development, the shock model has been studied extensively and enhanced in several ways. One of the first extensions is to permit successive shocks to cause increasing amounts of damage. That is, we continue to assume that the shocks occur according to a Poisson process and that the damage caused by the shocks is independent, but successive shocks are increasingly harmful, so the damage associated with the i^{th} shock, X_i, has distribution $G_{X_i}(x)$, with the values of the $G_{X_i}(x)$ decreasing in i. With this assumption, the survival function becomes

$$\bar{F}_T(t) = \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} G_{X_0} * G_{X_1} * \cdots G_{X_k}(L) \qquad (5.9)$$

where $G_{X_0} * G_{X_1} * \cdots G_{X_k}(L)$ denotes the convolution of the distributions.

Further extensions include (1) eliminating the assumption of independence of the successive shocks, (2) allowing attenuation of damage (partial healing) between shock events, and (3) treating the failure threshold as a decreasing function of time or damage. In each case, the basic model format is the same, and the resulting life distribution is IFRA.

The cumulative damage models have been viewed as very useful and have been applied to some very important problems. Perhaps the most noteworthy application was the study by Birnbaum and Saunders [24] of fatigue failures in aircraft fuselage. The development and use of such models continues.

5.2 ELECTRONIC FAILURE MODELS

5.2.1 The Arrhenius Model

To a large extent, models of electrical and electronic device reliability are motivated by empirical observations and were developed subsequent to the mechanical reliability models. The rapid evolution of electronic and especially microelectronic devices provided considerable impetus to the development of reliability models for electronic devices. In addition, because the life lengths of early electronic devices were often considerably shorter than those of mechanical devices, electronic device reliability received substantial attention.

Most of the models developed are based on the idea that electronic device degradation processes are essentially chemical conversion reactions within the materials that comprise the devices. Consequently, many of the models are based on the Arrhenius [31] reaction rate equation, which is named for the nineteenth-century chemist who developed the equation in the study of irreversible reactions such as oxidation. The basic form of the equation is that reaction rate, ρ, is given by

$$\rho = \eta e^{-E_a/KT} \tag{5.10}$$

where η is an electron frequency factor, K is Boltzmann's constant (8.623×10^{-5} ev/°K), T is temperature in degrees Kelvin, and E_a is the Gibb's free energy of activation. For reliability modeling, the product of the reaction rate, ρ, and time yields the extent of the progress of the deterioration reaction and is therefore considered to correspond to the cumulative hazard at any time. For example, for the Weibull life distribution

$$Z_T(t) = \left(\rho t / \theta \right)^{\beta} \tag{5.11}$$

and it may be further noted that alternate rate functions may be (and sometimes are) used but are applied in the same manner. Until recently, the scientific and algebraic links between rate functions such as Equation 5.10 and life distributions and hazard functions has not been widely studied. Nevertheless, the implied models such as Equation 5.11 are used extensively and have proven to be quite accurate.

5.2.2 The Eyring Model

A more general model was suggested by Krausz and Eyring [32]. In their model, the rate of progress of the deterioration reaction depends upon both temperature and voltage, so

$$\rho = \eta e^{-E_a/KT} e^{-\gamma_1 V - \gamma_2 V/KT} \tag{5.12}$$

where γ_1 and γ_2 also represent activation energies. Note that in this model, Krausz and Eyring allow for the possibility that, in addition to their direct effects, temperature and voltage have a synergetic (combination) effect on reaction rate. For this model, we again incorporate the rate function in the cumulative hazard function to obtain a life distribution.

5.2.3 Power Law Model

An alternate model for the effect of voltage is provided by the power law model. In this case, it is assumed that the mean life length is proportional to the (reciprocal of the) voltage at which a device is operated raised to an appropriate power. That is,

$$E[T] = c/V^p \tag{5.13}$$

Clearly, the two parameters of this model provide sufficient flexibility that any observed value of the mean life can be matched by the model.

5.2.4 A Defect Model

For the specific case of integrated circuits, Stevenson and Nachlas [33] suggest that metallization lines contain both macroscopic and microscopic flaws, and that these two flaw types both contribute to device failure. Their basic premise is that the flaws correspond to material impurities or crystal lattice anomalies that serve as reactive sites for degradation reactions such as metal migration and oxidation. Based on this basic premise, they show that published microelectronic failure data is reasonably well described by either the Weibull or the IDB life distributions, depending upon the relative magnitudes of the initial concentrations of the two defect types.

5.3 OTHER FAILURE MODELS

In recent years, some more useful and realistic, or at least more general, models have been proposed. Of these, the proportional hazards models [34] have been widely used, and the diffusion models have not yet been widely adopted but should be in the future. Another more general model that has been used in actuarial studies for a long time is the competing risk model. This model has not yet received much attention from the reliability community, but it soon will. These three model forms are examined below.

5.3.1 A Diffusion Process Model

The diffusion process models are really a generalization of the cumulative damage models. They seem to apply well to mechanical failure processes, but are certainly not limited to mechanical components. The models were best articulated by Lemoine and Wenocur [35]. They suggest that the accumulation of age or stress or the gradual reduction in strength of a component may be modeled as a diffusion process. In the absence of other phenomena, the diffusion process represents the deterioration of the component, and the first passage distribution of the process represents the failure time. They then add the possibility of sudden catastrophic events occurring in time according to an arbitrarily and selectable rule. In this case, the component ages gradually and fails according to the first passage time distribution unless the component encounters a sufficiently severe shock to cause it to fail "early." A familiar example is an automobile tire that gradually wears and ultimately is worn out. As it wears, the tire has reduced tread or strength and thus less resistance to randomly occurring road hazards, some (but not all) of which are fatal to the tire. A conceptually equivalent process is the gradual deterioration of an incandescent light bulb that will also fail suddenly if subjected to a voltage surge.

The general model that Lemoine and Wenocur define can be quite difficult to analyze or evaluate computationally. However, as they point out, the practically useful model realizations are generally tractable. Under their formulation, component status in time is represented by a state variable $X(t)$ for $t \geq 0$. The state variable evolves according to a diffusion process with drift $\mu(t)$ and diffusion coefficient $\sigma^2(t)$. Physically, $X(t)$ is the level of wear (reduction in strength) at time t, and it has mean value $\mu(t)$ and standard deviation $\sigma(t)$. Then, they define $k(x)$ to be the "killing function" for a component in state x (having accumulated wear x). The function $k(x)$ is essentially the rate at which shocks of sufficient magnitude to cause the component to fail occur when a component is in state x. In fact, this means that for most of the useful applications of the model, $k(x)$ is the hazard function.

For the diffusion-based model, the general Chapman-Kolmogorov backward differential equation is

$$\frac{\partial \bar{F}_T(x, t)}{\partial t} = -k(x)\bar{F}_T(x, t) + \mu(t)\frac{\partial \bar{F}_T(x, t)}{\partial x} + \frac{\sigma^2(t)}{2}\frac{\partial^2 \bar{F}_T(x, t)}{\partial x^2} \quad (5.14)$$

where the interpretation of $\bar{F}_T(x,t)$ is that it is the probability that a component having wear level x survives beyond time t.

This rather complicated model is actually quite rich. The choices of initial conditions and of functions $\mu(t)$ and $\sigma^2(t)$ lead to a wide variety of useful and conceptually reasonable specific models. For example, assuming $\sigma(t) = 0$ and that the wear tolerance is infinite implies that $k(x)$ corresponds to the Weibull hazard function. For that model, setting $\beta = 1$ to obtain an exponential distribution corresponds to the case in which the wear rate is zero. On the other hand, taking $k(x) = x$ and $X(t) = a + bt$ yields the Rayleigh distribution. Other choices yield Normal, Gamma and extreme value life distributions. Thus, as Lemoine and Wenocur point out, selecting the mean value and diffusion functions and the killing function to represent specific types of degradation processes leads to quite reasonable reliability models.

A generalization of the reaction rate view of electronic failure processes is suggested by Nachlas [30]. He treats irreversible deterioration reactions as pure birth processes in which the rate functions are the identifying characteristic of reaction type. Under this formulation, the fact that the birth process models are essentially specific cases of Poisson processes implies that all resulting life distributions are Gamma or generalized Gamma distributions. As a consequence, the models are probably more instructive than representative. They do have the appeal that they provide a more restricted but still the same type of physical interpretation as the Lemoine and Wenocur model.

5.3.2 Proportional Hazards

A relatively unified representation of the relationship between environment and failure processes is provided by the proportional hazards model that was originally proposed by Cox [36]. There are several forms in which this model may be stated. A general form that is conceptually appealing is

$$z_T(t) = \psi(\underline{x}) z_0(t) \tag{5.15}$$

The interpretation of the equation is that the device has a base hazard function, $z_0(t)$ that represents the core dispersion in life length, and this hazard is increased by a function, ψ, of the vector of variables, $\underline{x}$, that describe the specific operating environment in which the device is used.

There are actually two approaches to using the proportional hazards model. One is to identify the vector $\underline{x}$ as a set of *covariates* or *explanatory variables* and to view the model as a statistical basis for fitting a life distribution to observed failure data. An alternate approach is to view the vector $\underline{x}$ as a description of the operating environment and to treat the function ψ as a description of the effect of operating conditions on the failure frequency. Practically, the two views are essentially the same, because the function $\psi(\underline{x})$ is usually determined by regression analysis on accumulated failure data, and it is usually assumed that the function is linear. On the other hand, when the failure process for a device is understood, the proportional hazards model in which the function ψ represents that failure process is very appropriate.

Observe that, under the basic proportional hazards model, the shape of the hazard function is preserved. The environmental effects on the hazard are essentially additive, and the cumulative hazard function has the comparable form

$$Z_T(t) = \int_0^t \psi(\underline{x}) z_0(u)du = \psi(\underline{x}) \int_0^t z_0(u)du = \psi(\underline{x}) Z_0(t) \quad (5.16)$$

Thus, the reliability function is an exponential function of the environmental variables:

$$\bar{F}_T(t) = e^{-\psi(\underline{x}) Z_0(t)} \quad (5.17)$$

The advantages of this model are that it applies to any assumed base life distribution, it appears to be equally appropriate for mechanical and electronic devices, and it allows for the very specific representation of a general and arbitrary set of environmental effects. Conceptually, the chief drawback of the model is that it may not directly portray the mechanisms of failure. Practically, the drawback of the model is that substantial volumes of data are required to obtain satisfactory regression models for $\psi(\underline{x})$. Enhancements to the model have been based on allowing some of the environmental variables to be time dependent.

5.3.3 Competing Risks

Competing risk models have been used for actuarial studies for a very long time. Some people say their use started with Sir Thomas Bayes, when he attempted to compute the probabilities of death from various causes given that one cause, tuberculosis, was eliminated. In any case, the concept of the competing risk model is that a person or a device is subject to failure as a result of the action of multiple failure (or deterioration or disease) processes, which are competing to be the cause of failure (death). In the case of human disease processes, it is inappropriate to treat the various causes as independent, so the models are often quite complicated. This may be the reason that the application of the competing risk models to equipment reliability has been fairly limited.

The basic approach to constructing a competing risk model is to treat the various risks as components of a series system. The system fails at the time of the earliest component failure, which is the time that the first of the competing risks reaches completion. Thus, if a device is subject to failure from K causes, C_k, $k = 1, \ldots, K$, and if the time at which cause C_k reaches completion is t_k, then the life length of the device is

$$T = \min_k \{t_k\}$$

The analysis of the competing risk model starts with a consideration of the "net" probabilities, which are essentially the marginal probabilities for each risk — the life distribution assuming other risks are not present. Assuming there exists a joint life distribution, the associated joint survivor function is

$$\bar{F}_{T_1, T_2, \ldots, T_K}(t_1, t_2, \ldots\ldots, t_K) = \Pr[T_1 \geq t_1, T_2 \geq t_2, \ldots., T_K \geq t_K]$$

$$= \int_{t_1}^{\infty} \cdots \int_{t_K}^{\infty} f_{T_1, T_2, \ldots, T_K}(t_1, t_2, \ldots\ldots, t_K) dt_1 \cdots dt_K \qquad (5.18)$$

so the marginal (net) survival function is

$$\bar{F}_{T_j}(t_j) = \bar{F}_{T_1, T_2, \ldots, T_K}(0, 0, \ldots, 0, t_j, 0, \ldots .0) \qquad (5.19)$$

and the overall survivor function for any time, T, is

$$\bar{F}_T(t) = \bar{F}_{T_1, T_2, \ldots, T_K}(t, t, \ldots ., t) \qquad (5.20)$$

The probability that failure, when it occurs, is due to specific cause J is

$$\begin{aligned}\pi_j &= \int_0^\infty f_{T_j}(u)du \\ &= \int_0^\infty -\frac{\partial}{\partial t_j}\bar{F}_{T_1, T_2, ..., T_K}(t_1, t_2, ..., t_K)\Big|_{t_1=u, ..., t_K=u} du\end{aligned} \tag{5.21}$$

Next, one computes "crude" failure probabilities for each risk, which are the probabilities of failure due to each risk when other possible causes are considered. To construct the "crude" probabilities, one constructs the joint survival probability:

$$\Pr[T \geq t, J = j] = \Pr[T_j \geq t, T_j \leq T_k, \forall k \neq j]$$

$$= \int_t^\infty \left(\int_{t_j}^\infty \cdots \int_{t_j}^\infty f_{T_1, T_2, ..., T_K}(t_1, t_2, ..., t_K) \prod_{k \neq j} dt_k \right) dt_j \tag{5.22}$$

One then takes the overall survivor function of Expression 5.20 to be

$$\bar{F}_T(t) = \Pr[T \geq t] = \sum_{k=1}^{K} \Pr[T \geq t, J = k] \tag{5.23}$$

and the cause probability of Expression 5.21 to be

$$\pi_j = \Pr[J = j] = \Pr[T \geq 0, J = j] \tag{5.24}$$

Finally, the survivor function for the "crude" life distribution is

$$\bar{F}_{T|j}(t) = \Pr[T \geq t \mid J = j] = \frac{\Pr[T \geq t, J = j]}{\pi_j} \tag{5.25}$$

Observe that the described analysis applies regardless of whether or not the various risks are independent. Of course, the analysis depends upon having an initial multivariate model of life length as a function of the various risks. This presents a problem, because it is difficult to justify any particular multivariate model prior to observing failures, and when failures are observed, they correspond to observations of crude failures rather than net failures. The statistical analysis of such data is discussed in Chapter 8.

The appeal of the competing risk model is that many types of equipment really are subject to multiple competing risks. A further advantage of the model is that the combination of risks, even when each risk has a monotone hazard function, often yields a "crude" failure distribution that has a bathtub shape. Thus, the competing risk model provides a means to combine failure distributions to yield conceptually appealing model behavior in the overall hazard function.

Note finally that, like the proportional hazards model, the competing risk model applies to both mechanical and electronic devices and may be employed with any of the life distribution models. Thus, it is a model form with wide applicability.

5.4 EXERCISES

1. Assume that a component has a Weibull strength distribution with $\beta = 2.0$ and $\theta = 1267$ kg/cm^2, and that it is subject to a Normal stress distribution having $\mu = 105.5$ kg/cm^2 and $\sigma = 1.76$ kg/cm^2, where in both cases the distributions are time invariant. Compute the reliability of the component.
2. Resolve Problem 1 under the assumption that the stress distribution has increasing mean and variance according to the equations $\mu(t) = 105.5 + 0.0014t$ and $\sigma(t) = 1.76 + 0.0005t$.
3. Solve the cumulative damage model of Equation 5.6 for a component having a strength threshold $L = 60$ and subjected to shocks that impart a Normally distributed damage with mean $\mu = 17.5$ and standard deviation $\sigma = 1.5$. Assume shocks occur according to a Poisson process with rate $\lambda = 0.01$/hr.
4. Using the Arrhenius reaction rate model of Equation 5.8 and assuming a Weibull life distribution, compute the component reliability at 10,000, 25,000, and

50,000 hours when T = 55°C, $E_a = 0.80$, $\eta = 1.5 \times 10^{12}$, $\beta = 0.75$, and $\theta = 40{,}000$ hours. Then plot the reliability at 25,000 hours versus the activation energy for $0.5 \le E_a \le 1.2$.

5. Using the proportional hazards model, plot the baseline and overall hazard functions and the reliability function for a component having a Weibull life distribution with $\beta = 1.75$ and $\theta = 2500$ hours and $\psi(\underline{x}) = 1.4$.
6. Consider a competing risk model in which the number of risks is 2 and the joint density on life length is

$$f_T(t) = f_{T_1, T_2}(t_1, t_2) = 3e^{-t_1 - 3t_2}$$

Determine the probabilities of failure due to each cause, the net failure probability functions and the survival functions for each of the crude life length probabilities.

6

Age Acceleration

Models of failure processes and the effects of environmental conditions on failure behavior are often used as a basis for defining age acceleration regimes. As the term implies, these are methods for causing a device to age more rapidly than normal. The basic principle is that altering the operating environment so that it is harsher than normal operating conditions will increase the rate at which the device ages. That is, applying "stress" to the device will cause it to age more rapidly. The parallel idea of increased aging in humans due to excessive stress is considered representative.

Age acceleration of components or systems is generally used for one of two purposes — accelerated life testing or stress screening. In both cases, the environmental models are the same.

As the term implies, accelerated life testing is the use of age acceleration to study device longevity. The focus is on testing and the support of product design and development. Estimates of life distribution parameter values, or at least tolerance bounds on reliability, may be obtained. Unsatisfactory reliability performance may be recognized, or sensitivity to specific environmental factors may be investigated. The idea in accelerated life testing is that the time needed to test a new component design in order to verify its reliability can be excessive. In fact, many current durable good products such as computers and automobiles have components for which typical life lengths are measured in years. In principle, testing a new component design to verify that it is sufficiently reliable should involve observation of a sample of the devices operating until at least some of them fail. In the absence of an age acceleration method, the design verification effort would be unreasonably long. Age acceleration is used to reduce the test duration.

In contrast to accelerated life testing, stress screening is the use of age acceleration to age a device population beyond the early life phase of a product population so that the units put into customer use have reduced and presumably low hazard. It is a means to insulate the customers of a fixed product design from initially high hazard rates of DFR devices. This should reduce the level of expense for warranty support and also assure a favorable product reputation. In fact, optimization models have been defined for selecting the most cost-effective stress regimen. In these models, the costs of applying the environmental stress to the devices and the costs of lost new product are balanced against the avoidance of the costs of failures that would be experienced as a result of failures during customer use of the product. The solutions obtained indicate that, for many types of electronic devices, stress screening is a worthwhile strategy for managing device reliability.

Here again, there is an interesting contrast between mechanical and electronic devices. Experience has shown that many electrical and electronic devices display high and decreasing hazard during early life, so stress screening can be economically sensible. On the other hand, most mechanical components display increasing hazard from the start of life. Consequently, stress screening is rarely used for mechanical components, and models of its use have not been constructed.

For both applications, the operating environment of a device is modified within the limits that the device should be able to tolerate. That is, environment is manipulated in such a manner that the aging process remains unchanged except that it occurs more rapidly. It is important to maintain the nature of the failure processes, so that the observed aging behavior is consistent with actual use. When the age acceleration is done properly, it is usually believed that a proportional hazards type of hazard rate enhancement is obtained, so the shape of the distribution is maintained. Thus, age acceleration yields only time-scale compression and

$$F_{T_a}(t) = F_T(at) \tag{6.1}$$

where "a" is the acceleration factor, so Expression 6.1 states that the life distribution under enhanced stress is the same as the life distribution under nominal operating conditions but evaluated for a compressed time scale.

Consider the example of a component having a Weibull life distribution with $\beta = 1.5$ and $\theta = 1000$ hours. If the application of a stress to the component population results in an age acceleration of $a = 10$, the density and distribution functions are compressed as shown in Figure 6.1. Representative values of the distribution functions are

$$F_{T_a}(50) = F_T(500) = 0.298$$

$$F_{T_a}(100) = F_T(1000) = 0.632$$

and

$$F_{T_a}(200) = F_T(2000) = 0.941$$

Note the compression in the range of dispersion. A key point is that the shape of the distribution is unchanged, and this reflects the fact that the basic failure process is unchanged.

6.1 AGE ACCELERATION FOR ELECTRONIC DEVICES

The models used to represent age acceleration are essentially those for the effects of operating environment presented in the preceding chapter. In the case of electronic components, the application is direct. Over the past 50 or 60 years, age acceleration has been performed regularly on electrical and

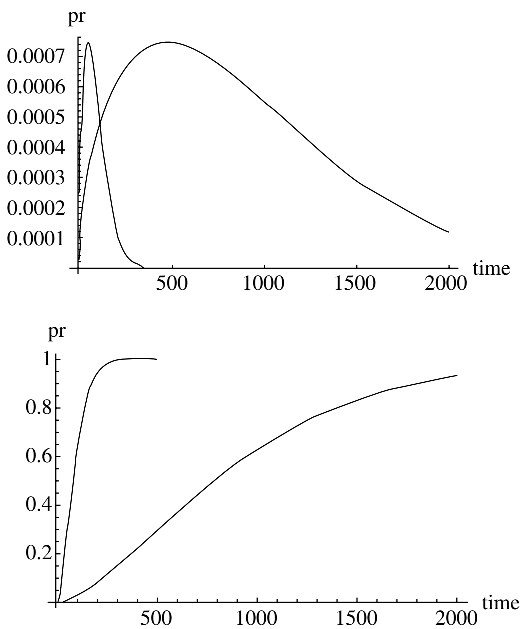


Figure 6.1 The density functions and the distribution functions with and without age acceleration.

electronic components. It may be the particular failure processes to which electrical devices are subject, or it may be their relatively short lives, but substantial evidence of the effectiveness of age acceleration has been accumulated. In fact, the extensive experience in the application of the Arrhenius, Eyring, and power law models has lead to general acceptance of their accuracy.

In the case of the Arrhenius model, the age acceleration factor "a" is computed as the ratio of the reaction rate under enhanced stress to that for the nominal conditions. Thus, if T_0 is the nominal operating temperature and T_a is the stress temperature, the age acceleration is

$$a = \frac{\rho(T_a)}{\rho(T_0)} = e^{\frac{E_a}{K}\left(\frac{1}{T_0} - \frac{1}{T_a}\right)} \tag{6.2}$$

As an example, consider a microelectronic device that is normally operated at a temperature of T_0 = 55°C = 328°K. If the device has an activation energy of E_a = 0.8 and is tested at T_a = 95°C = 368°K, it will age at a rate a = 21.6 times faster than under normal operating conditions. That is, for each hour of operation at 95°, the device gains 21.6 hours of age.

An interesting extension to this model was studied by Nachlas [37] in response to evidence presented by Jensen and Wong [38] that for CMOS devices, the activation energy is often temperature dependent. Recognizing that performing thermal age acceleration implies heating and cooling intervals, Nachlas defined a model in which the net age acceleration is obtained as a function of the heating and cooling rate and of the soak temperature. The basic form of that model is

$$a = \frac{e^{\frac{E_a(T_0)}{KT_0}}}{D} \int_0^D e^{-\frac{E_a(T(t))}{KT(t)}} dt \qquad (6.3)$$

where D is the test duration including heating and cooling intervals, the function $T(t)$ describes the temperature profile over time, and the function $E_a(T)$ depends upon the device identity and may be selected on the basis of the findings of Jensen and Wong. A representative temperature profile is shown in Figure 6.2. The cycle represented shows heating and cooling intervals of 200 minutes each with a dwell (or soak) period of 800 minutes. Representing the temperature profile by

$$T_a(t) = \begin{cases} 294 + 74\left(t/200\right)^{2.5} & 0 \le t \le 200 \\ 368 & 200 \le t \le 1000 \\ 368 - 74\left((t-1000)/200\right)^{2.5} & 1000 \le t \le 1200 \end{cases}$$

and on the basis of the Jensen and Wong paper, taking the activation energy function as

$$E_a(T_a) = 1.843e^{-0.00264T_a}$$

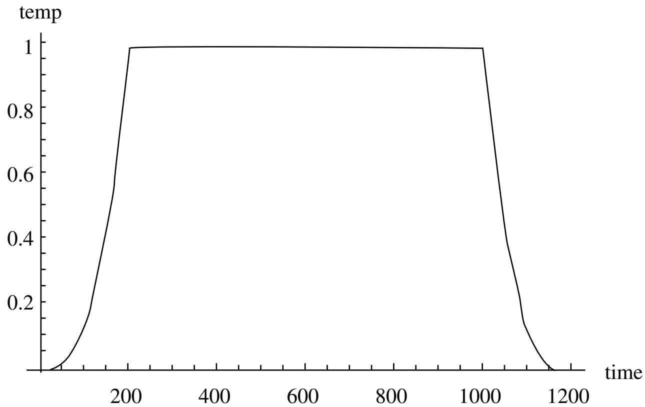


Figure 6.2 Representative thermal cycle for age acceleration.

numerical integration indicates an acceleration factor of a = 167.5.

For the Eyring model, assume the two temperatures are the same as in the above example and that the nominal and stress voltages are V_0 = 5.0 millivolts and V_a = 8.0 millivolts, respectively. The general acceleration factor equation is

$$a = \frac{\rho(T_a, V_a)}{\rho(T_0, V_0)} = e^{\frac{E_a}{K}\left(\frac{1}{T_0} - \frac{1}{T_a}\right)} e^{\gamma_1 (V_a - V_0)} e^{\frac{\gamma_2}{K}\left(\frac{V_a}{T_a} - \frac{V_0}{T_0}\right)} \tag{6.4}$$

so taking $E_a = 0.8$, $\gamma_1 = 0.2$, and $\gamma_2 = 0.008$, the age acceleration is a = 72.0.

Note that some types of components may not experience a synergistic effect of the combined stresses. In this case, $\gamma_2 = 0$, and the stresses are independent. They, nevertheless, both contribute to the acceleration in aging.

The (inverse) power law model is defined in terms of the expected life length. The age acceleration factor is therefore the ratio of the expected life under nominal conditions to the expected life length under increased stress. The expression is

$$a = \frac{E[T \mid V_0]}{E[T \mid V_a]} = \frac{c / V_0^p}{c / V_a^p} = \left(\frac{V_a}{V_0}\right)^p \tag{6.5}$$

In the case of the microelectronic device of the previous example, taking p = 2 implies an age acceleration of a = 2.56.

Two final points related to age acceleration of electronic devices are (1) that thermal stress is generally considered to be the most effective environmental modification for age acceleration, and (2) that Equations 6.1 and 6.2 each imply that the device hazard function may be represented in a proportional hazards format in which the temperature effect is a time-independent coefficient.

Note further that the evolution of modern computers, especially the laptop models, has been dominated by the development of more efficient and compact cooling systems.

6.2 AGE ACCELERATION FOR MECHANICAL DEVICES

In contrast to the case of the Arrhenius model for electronic devices, no general model has been defined for age acceleration of mechanical components. In fact, it is generally considered that, for each mechanical device, an age acceleration strategy, if there is one, must be identified specifically. Usually, this means evaluating the forces that drive the deterioration process and seeking a way to intensify those forces.

It is important to note that "stress" testing of mechanical devices is often misrepresented. For example, the hinge on the door of an automobile is typically used a few (less than 20) times per day. In advertising the durability of their automobiles, some firms show a door being opened and closed continuously and imply that this is a form of testing routinely used to verify reliability. While this demonstration may be impressive, and while the time until the hinge fails may be reduced, the increase in the frequency of use of the hinge does not constitute age acceleration. The forces imposed on the hinge during each open–close cycle are not changes, and thus the distribution on the number of cycles to failure is not compressed. The rate of actuation (or use) is increased, but aging is not accelerated. In the case of the hinge, age acceleration might be accomplished by increasing the weight of the door.

Another example of a test that is presented to the public is one in which an automotive tire is spun continuously and at relatively high speed on a test fixture. Here again, the usual assertion is that this test format accelerates aging and demonstrates the reliability of the product. The frequency of actuation is being used to reduce test time, but the distribution on the number of kilometers to failure is probably not being altered by the continuous use. It is true that increased turning speed might imply some age acceleration, and continuous use may imply heat accumulation that yields acceleration. These are worth studying, but the representation to the consumer public is misleading and inaccurate.

The guiding principle for age acceleration is, therefore, that the operating conditions experienced by a device must be modified in a manner that results in a compression in scale of the life distribution. Vassilou and Mettas [39] provide a simple but very informative example of this idea. They describe the "testing" of a paper clip by bending the (small) inner loop of the clip out of the plane that contains the entire unit. They consider that bending the inner loop out of the plane and back into position constitutes a single functional cycle and they define a "test" in which the usual angle of displacement of 45° is increased to the "stress" condition of 90°. An additional "test" at 180° is also possible. This paper clip example serves well to illustrate the fact that it is the modification of the degree to which the forces that drive failure that provides age acceleration. For mechanical devices, the failure processes must be determined, and the driving forces must be identified. Then a means to intensify those forces must be found. This process is usually specific to the component design.

The models presented in Chapter 5 do provide a format for the study of age acceleration of mechanical devices. In the case of the cumulative damage model of Equation 5.6, the frequency of activation effects can be represented easily by adjusting the rate parameter, λ, of the Poisson process for the arrival of shocks. While this does not comprise age acceleration, the adjustment permits the modeling of the revised operating profile. More useful models are obtained if the strength threshold parameter, L, is expressed as a function of the quantities that define operating conditions.

In a similar manner, both the stress-strength and proportional hazards models offer the potential to represent age acceleration. In the case of the stress-strength models, representing the parameters of the strength distribution in terms of operating conditions may provide a useful model. Normally, the proportional hazards models are based on variables (covariates) that represent the operating conditions, so the representation of age acceleration should be direct.

Finally, note that the diffusion process models have a "killing function," which represents the process of deterioration. Here again, if the model that comprises the "killing function" is defined (as would be expected) in terms of the forces that drive failure, the representation of age acceleration should again be direct.

6.3 STEP STRESS STRATEGIES

The effectiveness of age acceleration in testing the reliability of new device designs has lead to the development of numerous stress application strategies. One that has been found to be particularly useful is the step stress strategy. Under this plan, the key environmental stress variable, which is often temperature, is applied at successively higher levels. Thus, surviving devices are subjected to an increasingly aggressive environment. The failure data that is obtained in this manner will often support both the construction of a life distribution model for the device population and the estimation of failure model parameters such as the activation energy.

The number of levels at which the stress is applied depends upon the device being tested, the capability of the equipment used for stress application, and the extent of the information sought. Thus, the environmental profile is specific to each application. An example of a step stress strategy would be to test a sample of electronic devices having a constant activation energy of E_a = 0.70 at 95°C for 80 hours and to then increase the temperature to 110°C for an additional 40 hours. Under this regime, the devices that survive the entire test will experience an average age acceleration over the 120 hours of a = 21.47. A more informative measure is that devices that survive the test will have a net equivalent age of 2577 hours. The average age acceleration factor is computed as

$$a = \frac{\left(80e^{\frac{0.7}{0.00008623}\left(\frac{1}{328}-\frac{1}{368}\right)} + 40e^{\frac{0.7}{0.00008623}\left(\frac{1}{328}-\frac{1}{383}\right)}\right)}{120}$$

$$= \frac{\left(80(14.73)+40(34.96)\right)}{120} = 21.47$$

Of course, the failure data obtained during the test should provide considerable information about the failure characteristics of the device population.

One further example of a step stress test is provided by the paper clips. Suppose the stress regimen is to bend each member of a sample set of clips 20 cycles of 45° displacement, followed by 20 cycles of 90° displacement, and further followed by 20 cycles of 180° displacement. Clearly, the successive stress levels are increasingly aggressive and should produce increasing numbers of failures.

6.4 EXERCISES

1. Consider an electronic component having activation energy E_a = 0.6 and normal operating temperature T_o = 55°C. Compute and plot the age acceleration factor for the Arrhenius model that is realized for the range of temperatures 70° to 110°C.
2. Assuming a normal operating temperature of 55°C and an accelerated temperature of 95°C, compute and plot the acceleration factor defined by the Arrhenius model for E_a in the range of 0.4 to 0.9.

7

Nonparametric Statistical Methods

The specific description of actual failure characteristics of components is naturally based on the analysis of observed failure data. Sometimes, failure data is accumulated during the operation of a population of devices, but more often, the failure data is obtained through controlled testing of a sample of the devices. In either case, statistical methods are then used to obtain estimates for the parameters of life distribution models or to determine estimates of the reliability without considering specific models.

To a large extent, the statistical methods used to analyze reliability data are the same methods that have been defined for other statistical applications. In some cases, the methods require tailoring for application to failure data, but usually the application is direct. As suggested above, data may be analyzed to obtain estimates of the parameters of an assumed life distribution. The methods for computing estimates of distribution parameters are called "parametric" statistical methods. A variety of parametric methods are presented in the next chapter. Data analysis that is directed toward "model-free" estimates of reliability values and failure process characteristics are called "nonparametric" methods, and these methods are the subject of this chapter.

The chief advantage provided by nonparametric statistical methods for studying device reliability is that they do not require any assumption concerning the identity or the form of the life distribution. No potentially confining restrictions on behavior need be assumed. Also, for cases in which the choice of life distribution model is difficult, perhaps because the hazard function is not monotonic, the use of nonparametric methods permits the analyst to avoid the issue of model choice altogether. A second and also significant advantage of the nonparametric methods is that the quantities estimated are often easier to compute and manipulate than those derived from parametric methods. One further advantage of the nonparametric methods is that some permit the use of grouped data sets in which not all of the failure times are actually observed.

The key disadvantage of the use of nonparametric methods is that the estimates obtained usually do not support inferences about the identities of the hazard or distribution functions. In general, the quantities computed using nonparametric methods contain less information than those obtained with parametric analysis.

7.1 DATA SET NOTATION AND CENSORING

Before examining the specific analytical methods, it is worthwhile to establish some standard notation and some terminology that is descriptive of the process of collecting failure data. To define a standard notation, assume that (unless otherwise specified) the failure data that is to be analyzed has been obtained in a controlled test of a sample of "n" identical copies of a device. Assume further that each copy of the device that is tested is indexed so that its identity is known. We might think of the index as the "test slot" occupied by the device during the test.

We then take the set of observed failure times to be t_i, $i = 1, \ldots, n$. Thus, t_i is the time at which the component with index "i" fails. Consider an example. Suppose a sample of $n = 50$ copies of a device are tested and the (scaled) failure times are recorded as shown in Table 7.1. Associated with each copy of the device is an index, and for each indexed device, the time of failure is recorded.

Table 7.1 An Example Failure Data Set

i	t_i	i	t_i	i	t_i	i	t_i	i	t_i
1	0.883	11	1.555	21	0.129	31	0.829	41	0.894
2	0.875	12	3.503	22	0.455	32	0.548	42	0.336
3	5.292	13	1.541	23	2.008	33	1.016	43	0.129
4	0.038	14	1.218	24	0.783	34	0.223	44	1.373
5	4.631	15	1.285	25	1.803	35	3.354	45	0.613
6	1.690	16	2.190	26	2.505	36	1.559	46	1.272
7	0.615	17	0.720	27	0.465	37	3.785	47	0.019
8	2.877	18	0.056	28	1.494	38	0.599	48	0.068
9	1.943	19	0.006	29	0.795	39	0.090	49	0.658
10	3.106	20	0.279	30	0.299	40	0.026	50	3.085

Now, for many of the methods of analysis of the data, it is necessary that we reorder it so that the values are increasing. To represent the fact that the data has been reordered, we rename the values as x_j where

$$x_j = t_{[j]} \tag{7.1}$$

This expression should be read as x_j is the j^{th} smallest failure time observed, and it corresponds to the failure time of the device copy indexed [j]. For example, in the data set of Table 7.1, $x_1 = t_{[1]} = t_{19} = 0.006$, $x_2 = t_{[2]} = t_{47} = 0.019$, and $x_{50} = t_{[50]} = t_3 = 5.292$. The completely reordered data set is shown in Table 7.2. Now, the statistical methods that we consider will usually be applied to the values x_j.

With our notation defined, we next consider the question of whether or not a test is run to completion. It is easy to imagine a case in which the time required to test a complete sample of copies of a component until all have failed can be excessive. Even with age acceleration, the time required for a complete test can be unmanageable or infeasible. In order to perform a reliability test within a reasonable length of time, tests are often truncated early. A decision to use a truncated test should be made before the test is performed. When the test is truncated, we say that the test data obtained are censored.

Table 7.2 Reordered Example Failure Data

j	x_j	j	x_j	j	x_j	j	x_j	j	x_j
1	0.006	11	0.279	21	0.720	31	1.285	41	2.190
2	0.019	12	0.299	22	0.783	32	1.373	42	2.505
3	0.026	13	0.336	23	0.795	33	1.494	43	2.887
4	0.038	14	0.455	24	0.829	34	1.541	44	3.085
5	0.056	15	0.465	25	0.875	35	1.555	45	3.106
6	0.068	16	0.548	26	0.883	36	1.559	46	3.354
7	0.090	17	0.599	27	0.894	37	1.690	47	3.503
8	0.129	18	0.613	28	1.016	38	1.803	48	3.785
9	0.129	19	0.615	29	1.218	39	1.943	49	4.631
10	0.223	20	0.658	30	1.272	40	2.008	50	5.292

There are two basic approaches to test truncation — Type I and Type II. A test may be truncated at a preselected point in time or after a predetermined number of item failures. If the test is to be terminated after a fixed time interval, the test duration is known (and therefore limited) in advance, but the number of device failures that will be observed is a random variable. This is Type I test truncation, and the data set that results is said to be a Type I censored data set. On the other hand, truncation after a fixed number of failures yields a Type II censored data set, in which the number of data values is known in advance, but the test duration is random. In many cases, the distinction is not crucial, but for some statistical estimation methods, the difference between the two types of censoring is important. When a life test is not truncated, it is said to be a complete test, and the data set obtained is said to be a complete data set.

Now that our notation is established, and the basic classes of data sets that might be used are defined, we may examine the methods of analysis. To start, we consider only those cases in which a complete data set is available.

7.2 ESTIMATES BASED ON ORDER STATISTICS

A direct method for treating failure data is to use the number of observed failures or the number of surviving devices at any time to estimate the failure or survival probability. It seems intuitively reasonable to define a point estimate for the failure probability at any test time to be the ratio of the number of observed failures to the number of items on test. Similarly, the survivor probability can be estimated by the complement of the failure probability. That is, for any time t

$$\begin{aligned}\hat{F}_T(t) &= \frac{k}{n} \\ \hat{\bar{F}}_T(t) &= \frac{n-k}{n}\end{aligned} \qquad k = \{j \mid x_j \le t,\ x_{j+1} > t\} \tag{7.2}$$

where the caret, "^" above a quantity indicates an estimate. Now, while these expressions seem intuitively logical, they can be shown to be based on statistical reasoning and to be one of two reasonable forms.

Assume an arbitrary time interval, say $(t, t + \Delta T)$ and let p_j denote the probability that the j^{th} device failure observed occurs during that interval. In principle, any of the failures could occur during that interval so the probability that it is the j^{th} failure is a probability on the index of the failure time that happens to be the one in the interval of interest. Next, we ask how it could occur that the j^{th} failure would fall in the selected interval and the answer is that:

1. Each of the preceding $j - 1$ failures would have to occur before the start of the interval, and this occurs with probability $F_T(t)$ for each one.
2. Each of the succeeding $n - j$ failures must occur after the end of the interval, and this happens with probability $\bar{F}_T(t + \Delta t)$ for each one.
3. The j^{th} failure must fall within the interval, and this happens with probability $dF_T(t)$.
4. The number of ways the n copies of the device put on test can be separated into sets of $j - 1$, 1, and $n - j$ is given by the multinomial coefficient. Therefore

$$p_j = \frac{n!}{(j-1)!(1)!(n-j)!}\left(F_T(t)\right)^{j-1}\left(f_T(t)dt\right)\left(\bar{F}_T(t)\right)^{n-j} \tag{7.3}$$

This is the probability on the number of the failure that occurs during an arbitrary time interval and in the limit as the length of the interval is reduced, it is the probability that it is the j^{th} failure that has occurred by time t. Thus, it is the distribution on the "order statistic," which is the count of the number of failures. Properties of distributions on order statistics have been studied extensively. Among the results that are known and have been found useful are the facts that the mean of the distribution on the j^{th} order statistic implies that, as stated in Expression 7.2, an appropriate estimate for the fraction failed is

$$\hat{F}_T(t) = \frac{j}{n} \tag{7.4}$$

while the median of the distribution on the j^{th} order statistic implies that an appropriate estimate for the fraction failed is

$$\hat{F}_T(t) = \frac{j - 0.3}{n + 0.4} \tag{7.5}$$

Although it may seem counterintuitive, both of these estimates are reasonable, and each has advantages. The chief advantage of the estimator based on the mean, Expression 7.4, is that it is unbiased. That is

$$E[\hat{F}_T(t)] = F_T(t)$$

The estimator based on the median is not unbiased, but it has the advantage that it does not assign a value of 1.0 to the estimate associated with the n^{th} failure time. This is often considered important in that one does not expect that the greatest failure time observed during a test really corresponds to the maximum achievable life length. In general, either of the estimators may be used, and each is treated at various points in the discussion of statistical methods presented here.

7.3 ESTIMATES AND CONFIDENCE INTERVALS

The estimators defined in Equations 7.4 and 7.5 are called point estimates because they yield scalar values. From both a statistical and a design perspective, an interval estimate is usually more informative, because the point estimate is unlikely to exactly match the actual value of the estimated quantity. A confidence interval is a range for which there is an arbitrarily selected probability that the interval contains the true value of the estimated quantity. The construction of confidence intervals starts with the estimators and is described below.

It is also the case that the estimators defined in Expressions 7.4 and 7.5 may be viewed and analyzed in two related but rather different ways. Specifically, one may base the computation of the estimate of the reliability (or failure probability) on the number of survivors (or observed failures) or on the proportion of the test items that have survived (or failed). Consider first the use of the number of survivors.

The number of survivors at the end of a fixed time interval is a random variable for which the dispersion is best modeled using the binomial distribution. That is, survival of each copy of the component may be viewed as a Bernoulli trial, so that the quantity $n - j$ has a binomial distribution with success probability $\bar{F}_T(t)$. In this case

$$E[n-j] = n\bar{F}_T(t)$$

and

$$Var[n-j] = n\bar{F}_T(t)\left(1-\bar{F}_T(t)\right) \tag{7.6}$$

so the corresponding moments for the reliability are

$$E\left[\frac{n-j}{n}\right] = \bar{F}_T(t)$$

and

$$Var\left[\frac{n-j}{n}\right] = \frac{\bar{F}_T(t)\left(1-\bar{F}_T(t)\right)}{n} \tag{7.7}$$

In general, when the sample size is large (or at least not too small) and the success probability is not too close to 0 or to 1, a confidence interval for the binomial parameter may be computed using the standard normal distribution. For an arbitrarily selected confidence level of 100(1 − α)%, a confidence interval of

$$\hat{\bar{F}}_T(t) + z_{\alpha/2}\sqrt{\frac{\hat{\bar{F}}_T(t)\left(1-\hat{\bar{F}}_T(t)\right)}{n}} \le \bar{F}_T(t) \le \hat{\bar{F}}_T(t) + z_{1-\alpha/2}\sqrt{\frac{\hat{\bar{F}}_T(t)\left(1-\hat{\bar{F}}_T(t)\right)}{n}} \tag{7.8}$$

is valid asymptotically. Note that a table of standard Normal probabilities is provided in Appendix A along with numerical approximation strategies for Normal probabilities.

Consider an example. For the ordered data set provided in Table 7.2, the point estimates of the failure probability and of the reliability at t = 0.5 time units are

$$\hat{F}_T(0.5) = \frac{j}{n} = \frac{15}{50} = 0.30 \qquad \hat{\bar{F}}_T(0.5) = 1 - \bar{F}_T(0.5) = 0.70$$

$$\hat{F}_T(0.5) = \frac{j-0.3}{n+0.4} = \frac{14.7}{50.4} = 0.292 \qquad \hat{\bar{F}}_T(0.5) = 1 - \bar{F}_T(0.5) = 0.708$$

Taking α = 0.05, so that $z_{\alpha/2} = -1.96$, the first of the estimates provides the confidence interval

$$0.573 \le \bar{F}_T(0.5) \le 0.827$$

and the second one yields

$$0.582 \le \bar{F}_T(0.5) \le 0.834$$

A reasonably comparable but more involved estimate and confidence interval may be defined using the proportion of test items that have failed by any time. In general, the fraction failed, say $u = F_T(t)$, is best represented by the beta distribution for which the density function is

$$f_U(u, \eta, \delta) = \frac{\Gamma(\eta + \delta + 2)}{\Gamma(\eta + 1)\Gamma(\delta + 1)} u^{\eta}(1-u)^{\delta}\, du \tag{7.9}$$

In fact, this beta distribution form also follows directly from the distribution on the order statistics stated in Expression 7.3. With the definition $u = F_T(t)$ and the parameter identities of $\eta = j-1$ and $\delta = n-j$, replacing $f_T(t)dt$ by du yields Expression 7.9. Keep in mind that the proportion u must be in the interval (0, 1).

The point estimate of Expression 7.4 is an appropriate estimate for the fraction failed by time t, so a confidence bound on that estimate is obtained using the quantiles of the beta distribution. For integer values of the parameters, the distribution function for the beta distribution can be obtained by successive integration by parts and is

$$F_U(u, \eta, \delta) = 1 - \sum_{k=0}^{\eta} \frac{\Gamma(\eta + \delta + 2)}{\Gamma(k + 1)\Gamma(\eta + \delta - k + 2)} u^{k}(1-u)^{\eta+\delta-k+1} \tag{7.10}$$

Then, a 100(1 − α)% confidence interval on the failure probability at the time of the j^{th} observed failure is

$$u_{lower} \le F_T(t) \le u_{upper} \tag{7.11}$$

where u_{lower} is the solution to the equation

$$F_U(u_{lower}, j-1, n-j) = \alpha/2 \tag{7.12}$$

and u_{upper} is the solution to the equation

$$F_U(u_{lower}, j, n-j-1) = 1 - \alpha/2 \tag{7.13}$$

Clearly, these confidence limits must be computed numerically, but the effort required to do this is not great. For the same example case as the one treated above in which $n = 50$ and $j = 15$ at $t = 0.5$, the point estimate for $F_T(0.5)$ is still

$$\hat{F}_T(0.5) = \frac{j}{n} = \frac{15}{50} = 0.30$$

The corresponding 95% confidence interval is

$$0.175 \le F_T(0.5) \le 0.431$$

and for the reliability, this interval corresponds to

$$0.569 \le \bar{F}_T(0.5) \le 0.825$$

It is appropriate to reiterate the point that the above estimates and confidence intervals do not depend upon any assumed form for the life distribution. They apply to any set of test data, provided the data set is complete.

7.4 TOLERANCE BOUNDS

Returning to the beta density of Expression 7.9 for the proportion failed, let τ_γ represent the time at which the true cumulative failure probability is γ. That is

$$F_T(\tau_\gamma) = \gamma \tag{7.14}$$

Note that when γ is relatively small, the corresponding survival probability is $1 - \gamma$, and the quantity τ_γ is called the "reliable life." For example, $\tau_{0.01}$ is the age beyond which 99% of the population will fail. It is the time for which the reliability is 0.99. Similarly, $\tau_{0.10}$ is the age for which the reliability is 0.90. In an effort to standardize design practices and reliability demonstration for new component designs, the U.S. Air Force has identified, in its design guide [40], the 95% confidence limit on $\tau_{0.10}$ as a "Type A design allowable" and the 95% confidence limit on $\tau_{0.01}$ as a "Type B design allowable." The determination of these quantities is an important aspect of verifying the reliability of devices that are being proposed for Air Force use. In statistical terminology, these quantities are known as tolerance bounds.

The construction of tolerance bounds using test data is based on the distribution of order statistics and, hence, the beta density and distribution discussed above. To start, we consider the j^{th} failure time x_j, and as before we let

$$u = F_T(x_j)$$

Then for an arbitrarily selected value of γ, we wish to have a confidence level of 100(1 – α)% that τ_γ exceeds x_j. That is, we wish to have

$$\Pr[\tau_\gamma \geq x_j] \geq 1-\alpha$$

in which case, we will have 100(1 – α)% confidence that the reliability at x_j is at least γ. Using our established definitions, this means that the cumulative probability of failure at x_j is smaller than that at τ_γ, so

$$\begin{aligned}\Pr[\tau_\gamma \geq x_j] &= \Pr[F_T(x_j) \leq F_T(\tau_\gamma)] = \Pr[u \leq F_T(\tau_\gamma)] \\ &= F_U(F_T(\tau_\gamma),\ j-1,\ n-j) \geq 1-\alpha\end{aligned} \tag{7.15a}$$

Now, examining the beta distribution on a proportion u for the specific case of our test results, we find that:

$$\begin{aligned}F_U(u,\ j-1,\ n-j) &= 1-\sum_{k=0}^{j-1}\binom{n}{k}u^k(1-u)^{n-k} \\ &= 1-B(j-1,\ n,\ u)\end{aligned} \tag{7.15b}$$

where the notation $B(k, n, p)$ represents the cumulative binomial probability. Substituting this identity back into Equation 7.15a yields:

$$\begin{aligned}\Pr[\tau_\gamma \geq x_j] &= 1-B(j-1,\ n,\ F_T(\tau_\gamma)) \\ &= 1-B(j-1,\ n,\ \gamma) \geq 1-\alpha\end{aligned} \tag{7.15c}$$

so equivalently, the tolerance bound is defined by

$$B(j-1,\ n,\ \gamma) \leq \alpha \tag{7.15d}$$

Now, there are several ways in which we can use this result.

First, consider the cases associated with $j = 1$. For this case, Expressions 7.15b and 7.15d simplify to

$$F_U(u, 1, n-1) = 1-(1-u)^n = 1-(1-\gamma)^n \geq 1-\alpha$$

so we find that for the example data set of Table 7.2, using $\alpha = 0.05$ and $n = 50$, we compute $\gamma = 0.058$, and we say that we have 95% confidence that the reliability of the component population is at least $1 - \gamma = 0.942$ at a time of $x_1 = 0.006$.

An alternate use of Expression 7.15d is to ask how many items we should have tested in order for x_1 to correspond to a Type A design allowable. To answer this question, we set $\alpha = 0.05$ and $\gamma = 0.10$, and we compute n to be 29. For the corresponding case of the Type B design allowable, $\gamma = 0.10$ and $n = 298$.

Next, for other values of j, we might compute the reliability at the time of the 15th failure for which we have 95% confidence. To do this we solve Expression 7.15d for the smallest value of γ for which the expression holds. That value is $\gamma = 0.403$, so the reliability value we seek is 0.597, and we can say that

$$\Pr[\bar{F}_T(t = x_{15} = 0.465) \geq 0.597] \geq 0.95$$

Yet another possible computation would be to ask what level of confidence do we have that say $\bar{F}_T(t = x_{10} = 0.223) \geq 0.75$? The answer is

$$\Pr[\bar{F}_T(t = x_{10} = 0.223) \geq 0.75] \geq 0.738$$

Thus, using Expression 7.15d, we can calculate nearly any sort of tolerance bound we need, and we can also calculate the sample size required to obtain any specific level of confidence for any desired level of reliability.

7.5 TTT TRANSFORMS

A rather powerful nonparametric method for simply determining the behavior of the hazard function has been developed by Barlow and Campo [41] and explained very effectively by Klefsjo [42]. The method is known as the Total Time on Test (or TTT) transform. It has the appeal that, after some theoretical development, we obtain a simple graphical test. The underlying idea is that, for any life distribution (having finite mean), the relationship between the mean and the hazard function may be exploited to characterize the behavior of the distribution. In addition, the characterization can be performed for an empirical representation of the life distribution based on test data.

7.5.1 Theoretical Construction

In order to understand the behavior that is to be exploited, assume that a population of devices has a life distribution $F_T(t)$. Then, for any cumulative failure probability, say u, the inverse of the distribution function is denoted by $F_T^{-1}(u)$ and represents the time for which the cumulative probability of failure is u. Formally

$$F_T^{-1}(u) = \min\{t \mid F_T(t) \geq u\} \tag{7.16}$$

For example, if a population of devices has an exponential life distribution with λ = 0.84, then $F_T(0.75) = 0.467$, and $F_T^{-1}(0.467) = 0.75$. Now, using this notation, we have

> *Defn. 7.1*: The *Total Time on Test* (*TTT*) *transform* of a life distribution having finite mean, μ, is denoted by $H_{F_T^{-1}}(u)$ and is expressed as

$$H_{F_T^{-1}}(u) = \int_0^{F_T^{-1}(u)} \bar{F}_T(w)dw \tag{7.17}$$

Continuing with the above example of the exponential life distribution, the TTT transform value at $u = 0.467$ is

$$H_{F_T^{-1}}(u = 0.467) = \int_0^{F_T^{-1}(u=0.467)} \bar{F}_T(w)dw = \int_0^{0.75} e^{-\lambda w}dw$$

$$= \left(1 - e^{-0.75\lambda}\right)\Big/\lambda = 0.467\Big/\lambda = 0.556$$

The TTT transform has two useful and important properties that will lead us to a method for characterizing the behavior of a life distribution. The first is that it yields the mean when evaluated at u = 1.0. That is,

$$H_{F_T^{-1}}(u = 1.0) = \int_0^{F_T^{-1}(u=1.0)} \bar{F}_T(w)dw = \int_0^{\infty} \bar{F}_T(w)dw = \mu \quad (7.18)$$

To clarify, note that the time at which the life distribution has value 1.0 is infinity (or some very large maximum value), and the integral of the survivor function over the full range of any random variable yields the mean value of that random variable.

The second useful property of the transform is that its derivative, when evaluated at any value of the cumulative failure probability, equals the reciprocal of the corresponding value of the hazard function. This is shown as follows:

$$\frac{d}{du} H_{F_T^{-1}}(u) = \frac{d}{du} \int_0^{F_T^{-1}(u)} \bar{F}_T(w)dw$$

$$= \left(\frac{d}{du} F_T^{-1}(u) \right) \left(\bar{F}_T(F_T^{-1}(u)) \right) - \left(\frac{d}{du} 0 \right) \left(\bar{F}_T(0) \right) + \int_0^{F_T^{-1}(u)} \frac{d}{du} \bar{F}_T(w)dw$$

Now, clearly the second and third terms of this derivative equals zero. For the first term of the derivative, we observe that in the expression

$$\bar{F}_T(F_T^{-1}(u))$$

$F_T^{-1}(u)$ is the time for which the cumulative failure probability equals u, so evaluating the survivor function at that value yields $1 - u$. That is,

$$\bar{F}_T(F_T^{-1}(u)) = 1 - u$$

To evaluate the derivative of the inverse function, we proceed as follows:

$$t = F_T^{-1}(u)$$

so

$$\frac{dt}{du} = \frac{d}{du} F_T^{-1}(u)$$

Then

$$\frac{dt}{du} = 1 \Big/ \frac{du}{dt} = 1 \Big/ \frac{d}{dt} F_T(t) = 1 \big/ f(t)$$

and therefore,

$$\frac{d}{du} F_T^{-1}(u) = \frac{1}{f(F_T^{-1}(u))}$$

Combining our two results and evaluating the derivative at the cumulative failure probability associated with any time yields

$$\frac{d}{du} H_{F_T^{-1}}(u) = \left. \frac{1-u}{f(F_T^{-1}(u))} \right|_{u=F_T(t)} = \frac{\bar{F}_T(t)}{f(t)} = \frac{1}{z_T(t)} \qquad (7.19)$$

where the time indicated is that for which $t = F_T^{-1}(u)$.

The algebraic representations of the two properties of the TTT transform are combined below to obtain a characterization of the life distribution, but first we can scale the transform to make the ultimate test data analysis as generic as possible. We formally define a scaled TTT transform as:

> *Defn. 7.2*: The *scaled Total Time on Test (TTT) transform* of a life distribution having finite mean, μ, is denoted by $\Theta_{F_T^{-1}}(u)$ and is expressed as:

$$\Theta_{F_T^{-1}}(u) = H_{F_T^{-1}}(u) \Big/ H_{F_T^{-1}}(1) = H_{F_T^{-1}}(u) \Big/ \mu \qquad (7.20)$$

Now, clearly the derivative of the scaled transform equals the derivative of the transform divided by the scaling constant, μ. Thus,

$$\frac{d}{du}\Theta_{F_T^{-1}}(u) = \frac{\left(\frac{d}{du}H_{F_T^{-1}}(u)\right)}{\mu} = \frac{1/z_T(F_T^{-1}(u))}{\mu} \tag{7.21}$$

Consider what the derivative of the scaled transform tells us. Suppose the life distribution happens to be exponential. In that case, $z_T(t) = \lambda$, and $\mu = 1/\lambda$. Thus, at all values of u, the derivative of the scaled transform is

$$\frac{d}{du}\Theta_{F_T^{-1}}(u) = \frac{1/\lambda}{1/\lambda} = 1$$

When the life distribution has constant hazard, the scaled TTT transform has a slope equal to one. In that case, the scaled transform plots as a straight line. Suppose the hazard is not constant. For the Weibull distribution, the mean is $\theta\,\Gamma(1 + 1/\beta)$, and the hazard function is

$$z_T(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$$

For $\beta > 1$, the hazard function is increasing in time, and the scaled TTT transform displays the form shown in Figure 7.1a. Similarly, for $\beta < 1$, the scaled transform has the form shown in Figure 7.1b, as this is indicative of a distribution with decreasing hazard function. Thus, we have the very powerful result that the scaled transform is concave for increasing hazard, a straight line for constant hazard, and convex for decreasing hazard distributions.

7.5.2 Application to Complete Data Sets

The application of our understanding of the behavior of the scaled TTT transform is to plot the transform values computed

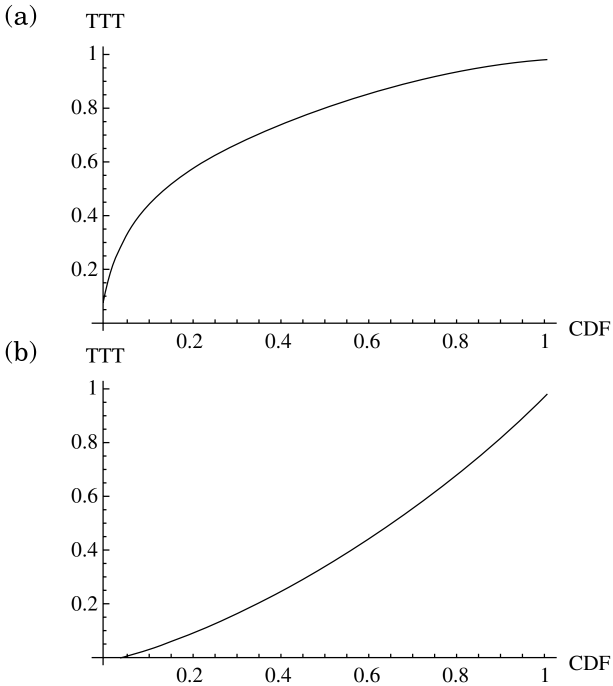


Figure 7.1 (a) TTT transform for increasing hazard functions. (b) TTT transform for decreasing hazard functions.

from test data to determine the likely form of the hazard function for the device being tested. To construct the application, we first define the quantities $\tau(x_j)$ as

$$\tau(x_j) = \sum_{k=1}^{j} x_k + (n-j)x_j \tag{7.22}$$

where the x_j are our ordered failure times, so $\tau(x_j)$ is the total amount of testing time that is accumulated by the time of the j^{th} failure. This may be seen as follows:

during the interval	the number of items on test is	so the test time accrued is
$(0, x_1)$	n	$n\,x_1$
(x_1, x_2)	$n-1$	$(n-1)(x_2-x_1)$
(x_2, x_3)	$n-2$	$(n-2)(x_3-x_2)$
•	•	•
•	•	•
•	•	•
(x_{j-1}, x_j)	$n-j+1$	$(n-j+1)(x_j-x_{j-1})$
•	•	•
•	•	•
•	•	•
(x_{n-1}, x_n)	1	(x_n-x_{n-1})

Taking the sum after any failure time, one has:

$$\tau(x_j) = \sum_{k=1}^{j} (n-k+1)(x_k - x_{k-1})$$

and this reduces to the form shown in Equation 7.22.

To apply the TTT transform to test data, we first construct the empirical life distribution corresponding to the test observations. For a sample of n items placed on test, let $F_{X_n}(t)$ represent the empirical life distribution. Thus,

$$F_{X_n}(t) = \begin{cases} 0 & 0 \le t < x_1 \\ 1/n & x_1 \le t < x_2 \\ \vdots & \vdots \\ j/n & x_j \le t < x_{j+1} \\ \vdots & \vdots \\ 1 & x_n \le t < \infty \end{cases} \tag{7.23}$$

Given this definition, the inverse function is defined by

$$F_{X_n}^{-1}(u) = \min\{x_j \mid F_{X_n}(x_j) \ge u\} \tag{7.24}$$

which is the smallest failure time with estimated cumulative probability equal to or greater than u. For example, for the data set shown in Table 7.2, we have

$$F_{X_n}(t)=\begin{cases} 0 & 0 \le t < 0.006 \\ 0.02 & 0.006 \le t < 0.019 \\ 0.04 & 0.019 \le t < 0.026 \\ 0.06 & 0.026 \le t < 0.038 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0.98 & 4.631 \le t < 5.292 \\ 1.00 & 5.292 \le t < \infty \end{cases}$$

A few values of the corresponding inverse function are $F_{X_n}^{-1}(0.04)=0.019$, $F_{X_n}^{-1}(0.30)=0.465$, and $F_{X_n}^{-1}(0.99)=4.631$. For the same data set, note that $\tau(x_1)$ = 50(0.006) = 0.30, $\tau(x_2)$ = 50(0.006) + 49(0.013) = 0.937, and $\tau(x_3)$ = 50(0.006) + 49(0.013) + 48(0.007) = 1.273.

The use of the TTT transform to analyze test data proceeds by applying the transform to $F_{X_n}(t)$ in the same manner as for $F_T(t)$. That is,

$$H_{F_{X_n}^{-1}}(u)=\int_0^{F_{X_n}^{-1}(u)} \bar{F}_{X_n}(w)dw \tag{7.25}$$

However, recognizing that $F_{X_n}(t)$ is a step function, the integral may be expressed as a sum. Consider the transform evaluated at $u = j/n$.

$$H_{F_{X_n}^{-1}}(j/n)=\int_0^{F_{X_n}^{-1}(j/n)} \bar{F}_{X_n}(w)dw=\int_0^{F_{X_n}^{-1}(1/n)} \bar{F}_{X_n}(w)dw$$
$$+\int_{F_{X_n}^{-1}(1/n)}^{F_{X_n}^{-1}(2/n)} \bar{F}_{X_n}(w)dw+\cdots+\int_{F_{X_n}^{-1}((j-1)/n)}^{F_{X_n}^{-1}(j/n)} \bar{F}_{X_n}(w)dw$$

$$= \int_0^{x_1} \frac{n}{n} dw + \int_{x_1}^{x_2} \frac{n-1}{n} dw + \cdots + \int_{x_{j-1}}^{x_j} \frac{n-j+1}{n} dw$$

$$= \frac{1}{n} \begin{pmatrix} nx_1 + (n-1)(x_2 - x_1) + (n-2)(x_3 - x_2) \\ + \cdots + (n-j+1)(x_j - x_{j-1}) \end{pmatrix}$$

$$= \frac{1}{n} \tau(x_j)$$

Thus, the TTT transform applied to the test data reduces to the sum of test times — the total testing time — defined in Equation 7.22. In addition, in the limit, the empirical transform corresponds to the theoretical transform for the underlying life distribution. That is

$$\lim_{\substack{n \to \infty \\ j/n \to s}} H_{F_{X_n}^{-1}}(j/n) = \lim_{\substack{n \to \infty \\ j/n \to s}} \int_0^{F_{X_n}^{-1}(j/n)} \bar{F}_{X_n}(w) dw$$

$$= \int_0^{F_T(s)} \bar{F}_T(w) dw = H_{F_T^{-1}}(s)$$

so the transform based on the empirical distribution is representative of the transform for the actual life distribution. Consequently, we can use the TTT transform obtained from the data to characterize the underlying life distribution. We actually use the scaled TTT transform. For the test data, the above limit also applies to the case of $s = 1$, so the scaling constant is $\tau(x_n)$. Thus, the scaled transform is

$$\Theta_{F_T^{-1}}(j/n) = \frac{H_{F_{X_n}^{-1}}(j/n)}{H_{F_{X_n}^{-1}}(1)} = \frac{\frac{1}{n}\tau(x_j)}{\frac{1}{n}\tau(x_n)} = \tau(x_j)/\tau(x_n) \tag{7.26}$$

The application to sample data is implied by the above discussion, and the result is that the quantities $\tau(x_j)$ and $\tau(x_n)$ are computed, and their ratio is plotted against j/n. If the

Table 7.3 Empirical Values of the Scaled TTT Transform

x_j	j/n	$\Theta(x_j)$	x_j	j/n	$\Theta(x_j)$	x_j	j/n	$\Theta(x_j)$
0.006	0.02	0.0046	0.613	0.36	0.3665	1.555	0.70	0.6951
0.019	0.04	0.0143	0.615	0.38	0.3674	1.559	0.72	0.6960
0.026	0.06	0.0194	0.658	0.40	0.3878	1.690	0.74	0.7240
0.038	0.08	0.0281	0.720	0.42	0.4162	1.803	0.76	0.7464
0.056	0.10	0.0407	0.783	0.44	0.4441	1.943	0.78	0.7721
0.068	0.12	0.0489	0.795	0.46	0.4492	2.008	0.80	0.7830
0.090	0.14	0.0637	0.829	0.48	0.4633	2.190	0.82	0.8108
0.129	0.16	0.0893	0.875	0.50	0.4815	2.505	0.84	0.8541
0.129	0.18	0.0893	0.883	0.52	0.4846	2.887	0.86	0.8947
0.223	0.20	0.1482	0.894	0.54	0.4886	3.085	0.88	0.9212
0.279	0.22	0.1824	1.016	0.56	0.5315	3.106	0.90	0.9231
0.299	0.24	0.1943	1.218	0.58	0.5993	3.354	0.92	0.9420
0.336	0.26	0.2158	1.272	0.60	0.6167	3.503	0.94	0.9511
0.455	0.28	0.2831	1.285	0.62	0.6206	3.785	0.96	0.9641
0.465	0.30	0.2886	1.373	0.64	0.6462	4.631	0.98	0.9899
0.548	0.32	0.3329	1.494	0.66	0.6794	5.292	1.00	1.00
0.599	0.34	0.3594	1.541	0.68	0.6916			

result is approximately a 45° line, one concludes that the life distribution is constant hazard. On the other hand, if the plot is concave and lies mostly above the 45° line, one concludes that the life distribution is IFR, and if the plot is convex and lies below the 45° line, one concludes that the life distribution is DFR.

Consider the example provided by the data of Table 7.2. The computed values of the scaled transform are shown in Table 7.3. Figure 7.2 shows the plot of the values in the table. Note that the points appear to resemble a 45° line. In fact, the plot seems to cross the 45° line several times and to generally lie close to it. An examination of the values in the table confirms this. It seems reasonable to conclude that the test data corresponds to a constant hazard life distribution.

In order to illustrate the contrasting behavior to constant hazard, consider the set of ordered test data in Table 7.4.

For this data, set the values of the scaled transform are listed in Table 7.5. The corresponding scaled TTT plot for the data is shown in Figure 7.3. Note that, except for some crossing

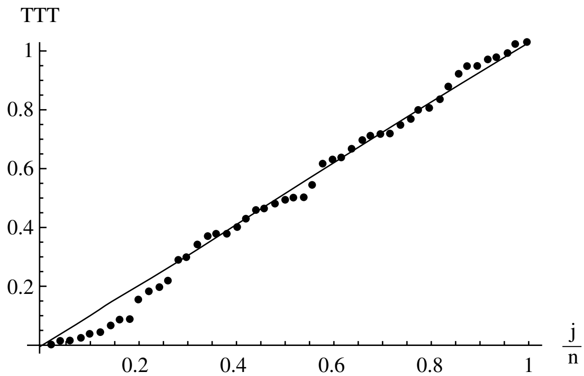


Figure 7.2 Plot of scaled TTT transform.

Table 7.4 Example Life Test Data Set

j	x_j	j	x_j	j	x_j	j	x_j	j	x_j
1	0.023	11	0.303	21	0.511	31	0.754	41	1.224
2	0.025	12	0.370	22	0.522	32	0.767	42	1.252
3	0.081	13	0.371	23	0.532	33	0.795	43	1.344
4	0.110	14	0.373	24	0.571	34	0.802	44	1.378
5	0.185	15	0.394	25	0.579	35	0.873	45	1.562
6	0.226	16	0.400	26	0.596	36	0.884	46	1.580
7	0.230	17	0.412	27	0.605	37	0.936	47	1.653
8	0.278	18	0.435	28	0.627	38	0.993	48	1.659
9	0.278	19	0.449	29	0.673	39	1.001	49	1.764
10	0.287	20	0.494	30	0.753	40	1.087	50	2.520

near zero and near 1.0, the entire plot lies above the 45° line. The appropriate interpretation is that the devices that produced the test data appear to display increasing hazard. The observed behavior is confirmed by an inspection of the transform values, which lie well above the corresponding values of j/n.

Both of the data analyses displayed so far used the entire complete data set. As discussed earlier, it is often impractical to run a test until all test units have failed. When the test is terminated early and a censored data set is obtained, it is still possible to use the TTT transform. The adjustments necessary to do this are treated in the following section.

Table 7.5 Empirical Values of the Scaled TTT Transform

x_j	j/n	$\Theta(x_j)$	x_j	j/n	$\Theta(x_j)$	x_j	j/n	$\Theta(x_j)$
0.023	0.02	0.0315	0.435	0.36	0.5121	0.873	0.70	0.7880
0.025	0.04	0.0342	0.449	0.38	0.5243	0.884	0.72	0.7925
0.081	0.06	0.1078	0.494	0.40	0.5625	0.936	0.74	0.8125
0.110	0.08	0.1451	0.511	0.42	0.5765	0.993	0.76	0.8328
0.185	0.10	0.2396	0.522	0.44	0.5852	1.001	0.78	0.8354
0.226	0.12	0.2901	0.532	0.46	0.5929	1.087	0.80	0.8613
0.230	0.14	0.2949	0.571	0.48	0.6217	1.224	0.82	0.8988
0.278	0.16	0.3514	0.579	0.50	0.6274	1.252	0.84	0.9057
0.278	0.18	0.3514	0.596	0.52	0.6391	1.344	0.86	0.9259
0.287	0.20	0.3615	0.605	0.54	0.6450	1.378	0.88	0.9324
0.303	0.22	0.3790	0.627	0.56	0.6588	1.562	0.90	0.9626
0.370	0.24	0.4506	0.673	0.58	0.6865	1.580	0.92	0.9651
0.371	0.26	0.4516	0.753	0.60	0.7325	1.653	0.94	0.9731
0.373	0.28	0.4537	0.754	0.62	0.7331	1.659	0.96	0.9735
0.394	0.30	0.4744	0.767	0.64	0.7398	1.764	0.98	0.9793
0.400	0.32	0.4801	0.795	0.66	0.7536	2.520	1.00	1.00
0.412	0.34	0.4913	0.802	0.68	0.7569			

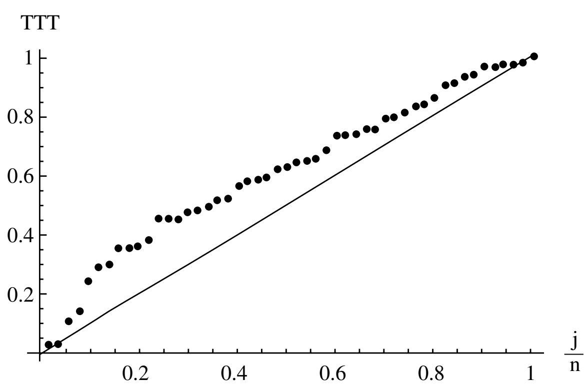


Figure 7.3 Plot of scaled TTT transform for an IFR distribution.

7.5.3 Application to Censored Data Sets

Suppose a life test has been performed on a sample of n = 50 items, and that the test was terminated early, so that r of the n items were observed to fail and $n - r$ had not yet failed. For

Table 7.6 Empirical Values of the Scaled TTT Transform for Censored Data Sets

		$r = 12$		$r = 20$
x_j	j/r	$\Theta(x_j)$	j/r	$\Theta(x_j)$
0.006	0.083	0.0236	0.05	0.0118
0.019	0.167	0.0736	0.10	0.0369
0.026	0.250	0.1000	0.15	0.0501
0.038	0.333	0.1444	0.20	0.0723
0.056	0.417	0.2094	0.25	0.1050
0.068	0.500	0.2519	0.30	0.1262
0.090	0.583	0.3280	0.35	0.1643
0.129	0.667	0.4598	0.40	0.2304
0.129	0.750	0.4598	0.45	0.2304
0.223	0.833	0.7627	0.50	0.3822
0.279	0.917	0.9387	0.55	0.4704
0.299	1.00	1.00	0.60	0.5011
0.336			0.65	0.5565
0.455			0.70	0.7299
0.465			0.75	0.7441
0.548			0.80	0.8585
0.599			0.85	0.9268
0.613			0.90	0.9450
0.615			0.95	0.9475
0.658			1.00	1.00

purposes of illustration, assume that only the first r data values of Table 7.2 had been recorded. In that case, the total observed test time is

$$\tau(x_r) = \sum_{k=1}^{r} x_k + (n-j)x_r \tag{7.27}$$

For the censored data set, this quantity is used in place of $\tau(x_n)$ as the scaling constant, and the scaled transform is defined as

$$\Theta_{F_T^{-1}}\left(j/r\right) = \frac{\frac{1}{n}\tau(x_j)}{\frac{1}{n}\tau(x_r)} = \tau(x_j)/\tau(x_r) \tag{7.28}$$

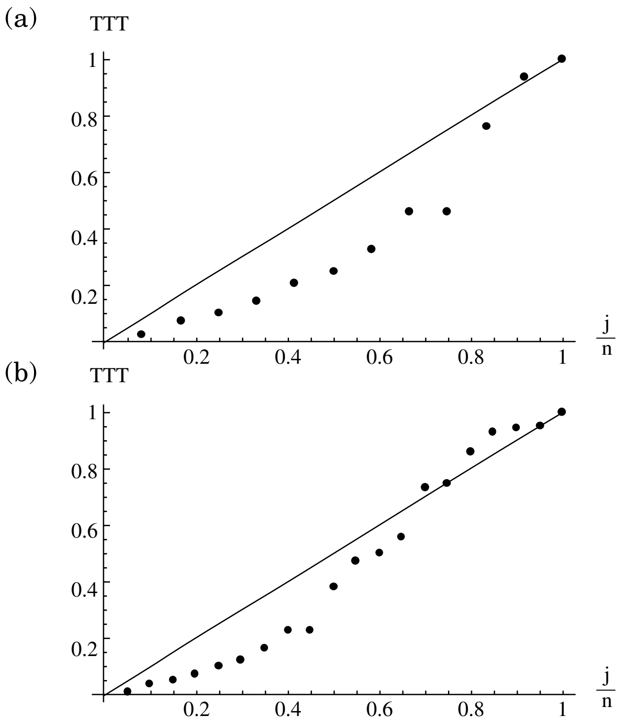


Figure 7.4 (a) Plot of scaled TTT transform for censored data with $r = 12$. (b) Plot of scaled TTT transform for censored data with $r = 20$.

and this quantity is plotted vs. j/r. For the example data set of Table 7.2, the computed values for r = 12 and for r = 20 are listed in Table 7.6. The corresponding plots of the scaled transform are shown in Figure 7.4a and Figure 7.4b, respectively. These plots serve to illustrate the facts that the transform may be applied to censored data and that our ability to interpret the plots is significantly influenced by the degree of censoring. For the case in which $r = 20$, we can be reasonably confident that the hazard is constant. For the plot corresponding to the data censored after $r = 12$ observations, our conclusion of constant hazard is rather more tenuous.

Table 7.7 Empirical Values of the Scaled TTT Transform for Censored Data from an IFR Distribution

	r = 12		r = 20	
x_j	j/r	$\Theta(x_j)$	j/r	$\Theta(x_j)$
0.023	0.083	0.0699	0.05	0.0560
0.025	0.167	0.0759	0.10	0.0608
0.081	0.250	0.2392	0.15	0.1916
0.110	0.333	0.3220	0.20	0.2580
0.185	0.417	0.5317	0.25	0.4260
0.226	0.500	0.6438	0.30	0.5157
0.230	0.583	0.6545	0.35	0.5243
0.278	0.667	0.7798	0.40	0.6247
0.278	0.750	0.7798	0.45	0.6247
0.287	0.833	0.8023	0.50	0.6427
0.303	0.917	0.8411	0.55	0.6738
0.370	1.00	1.00	0.60	0.8011
0.371			0.65	0.8028
0.373			0.70	0.8066
0.394			0.75	0.8434
0.400			0.80	0.8535
0.412			0.85	0.8734
0.435			0.90	0.9104
0.449			0.95	0.9321
0.494			1.00	1.00

Once again, to provide a contrast to the constant hazard case, consider the data from an IFR distribution that is listed in Table 7.4. If that data had been generated in a test with censoring either at r = 12 or r = 20, the corresponding data values would have been those shown in Table 7.7, and the corresponding plots of the scaled transforms would have been those shown in Figure 7.5a and Figure 7.5b. Here again, it is clear that the degree of censoring affects the confidence we have in our interpretations of the plots.

In closing this discussion, it should now be reasonably clear that the TTT transform provides a method that is very simple to perform for characterizing the hazard behavior of a device population.

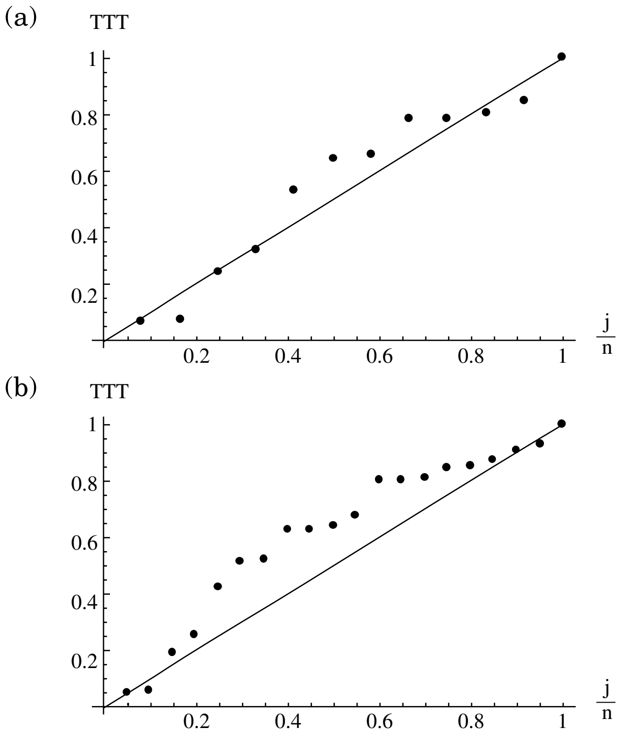


Figure 7.5 (a) Plot of scaled TTT transform for censored data with $r = 12$. (b) Plot of scaled TTT transform for censored data with $r = 20$.

It should also be apparent that nonparametric methods can provide us with substantial information concerning the reliability of a device design without requiring us to assume an underlying distribution model of the dispersion in failure times.

7.6 EXERCISES

1. The following data set was obtained from a life test of n = 50 copies of a component.

i	t_i	i	t_i	i	t_i	i	t_i	i	t_i
1	551.881	11	297.883	21	122.750	31	539.933	41	141.582
2	964.448	12	78.966	22	119.677	32	175.578	42	329.841
3	687.943	13	526.061	23	568.533	33	465.506	43	570.971
4	206.215	14	558.106	24	453.852	34	208.198	44	929.433
5	844.059	15	484.969	25	267.140	35	326.713	45	67.964
6	439.283	16	282.293	26	128.874	36	154.290	46	294.060
7	170.110	17	589.303	27	675.259	37	703.458	47	23.774
8	273.522	18	1032.227	28	347.812	38	327.022	48	295.930
9	475.883	19	726.202	29	283.398	39	511.423	49	514.202
10	255.646	20	573.447	30	357.552	40	560.902	50	251.874

   For this data, identify the values of x_1, x_5, x_{10}, and x_{50}.
2. For the data set of Problem 1, compute the point estimates of the reliability at x_1, x_5, x_{10}, and x_{50} with both the mean and the median based estimation equations. How do these estimates compare?
3. Using the data in Problem 1, compute 95% confidence intervals for the failure probability at each of x_1, x_5, x_{10}, and x_{50} using each of Expressions 7.8 and 7.11. How do these intervals compare?
4. Using the data of Problem 1, compute a 95% tolerance bound on the reliability at x_{10} and x_{15}. Then compute the level of confidence the data provides that the reliability at x_{20} exceeds 0.55.
5. For the following data set obtained from a life test, order the data, and then plot the TTT transform. Indicate what form of the hazard function is suggested by the plot.

i	t_i	i	t_i	i	t_i	i	t_i	i	t_i
1	635.655	11	456.731	21	335.464	31	282.015	41	170.998
2	369.012	12	459.482	22	100.790	32	172.954	42	465.023
3	312.489	13	420.944	23	453.539	33	216.222	43	254.582
4	196.092	14	306.064	24	82.843	34	204.064	44	319.789
5	72.393	15	216.330	25	356.053	35	228.195	45	285.048
6	22.150	16	180.638	26	255.021	36	528.971	46	307.34
7	302.257	17	137.704	27	302.217	37	270.25	47	318.541
8	114.434	18	159.855	28	181.568	38	117.524	48	242.783
9	68.381	19	231.442	29	93.694	39	70.280	49	458.005
10	200.899	20	203.094	30	314.594	40	93.151	50	130.900

6. For the data set of Problem 5, assume that because of test termination, only the earliest 16 data values are available. Construct the plot of the TTT transform for the resulting censored data set.
7. For the data set of Problem 5, assume that because of test termination, only the earliest 25 data values are available. Construct the plot of the TTT transform for the resulting censored data set.
8. Use the complete set of the following life test data to construct a plot of the TTT transform and indicate what type of behavior the hazard function appears to have.

i	t_i	i	t_i	i	t_i	i	t_i
1	29.835	11	1048.13	21	126.097	31	154.884
2	1262.860	12	641.953	22	434.761	32	103.444
3	804.623	13	762.882	23	170.046	33	176.225
4	691.363	14	206.062	24	1880.470	34	252.424
5	654.951	15	593.040	25	1058.727	35	333.961
6	409.087	16	224.793	26	957.271	36	1989.75
7	1615.690	17	203.809	27	2.970	37	1646.63
8	470.408	18	309.879	28	75.239	38	344.135
9	918.823	19	3094.740	29	346.605	39	48.831
10	68.348	20	55.854	30	801.645	40	131.215

9. For the data set of Problem 8, assume that because of test termination, only the earliest 12 data values are available. Construct the plot of the TTT transform for the resulting censored data set.
10. For the data set of Problem 8, assume that, because of test termination, only the earliest 20 data values are available. Construct the plot of the TTT transform for the resulting censored data set.

8

Parametric Statistical Methods

Parametric statistical methods for analyzing reliability data start with an assumption of the form of the life distribution. Usually, the choice of a distribution model is based on experience with similar types of devices or an understanding of the phenomena that determine item failure. In some cases, a general model such as the Weibull or Gamma is assumed, because the availability of two distribution parameters makes it likely that a reasonable representation of the failure probabilities will be obtained.

Once a life distribution is assumed, the statistical methods are used to obtain estimates for the parameters of those distributions — hence the terminology parametric methods. There are a large number of parametric reliability estimation methods. Only the principal methods are presented here. As each of the methods is discussed, it is illustrated by application to various life distributions with particular emphasis on the Weibull and exponential life distributions. The three principal methods that are discussed here are (1) the graphical method, (2) the method of moments, and (3) the likelihood method. Whenever necessary, the notational conventions discussed at the start of the last chapter are used here as well.

8.1 GRAPHICAL METHODS

Most of the graphical methods are based on the general relationship between the cumulative hazard function and the reliability function. As we know,

$$\bar{F}_T(t) = e^{-Z_T(t)} \tag{8.1}$$

so consequently,

$$-\ln \bar{F}_T(t) = Z_T(t) \tag{8.2}$$

If test data is used to compute estimated values of the reliability and these estimates are plotted as a function of time, the resulting graph will provide estimated values for the distribution parameters that form the hazard function.

To be more specific, assume the life lengths of a population of components are believed to be well represented by an exponential distribution. The realization of Equation 8.2 for the exponential is

$$\frac{1}{\ln \bar{F}_T(t)} = -\ln \bar{F}_T(t) = \lambda t \tag{8.3}$$

which is the equation for a line. Hence, if we represent our successive reliability estimates by

$$y_j = \ln \frac{1}{\hat{\bar{F}}_T(x_j)} = \ln \frac{n}{(n-j)} \tag{8.4}$$

then a plot of y_j versus x_j should be a line with slope of λ. Of course, while we can physically plot the observed data, we would be more likely to use linear regression to determine the line that best fits the data.

Table 8.1 Example Ordered Failure Data

j	x_j	$(n-j)/n$	y_j	j	x_j	$(n-j)/n$	y_j
1	4.740	0.950	0.051	11	128.756	0.450	0.799
2	12.636	0.900	0.105	12	150.393	0.400	0.916
3	17.358	0.850	0.163	13	168.101	0.350	1.050
4	22.099	0.800	0.223	14	194.277	0.300	1.204
5	29.085	0.750	0.288	15	238.897	0.250	1.386
6	32.732	0.700	0.357	16	303.383	0.200	1.609
7	41.725	0.650	0.431	17	340.621	0.150	1.897
8	57.518	0.600	0.511	18	382.142	0.100	2.303
9	62.864	0.550	0.598	19	492.023	0.050	2.996
10	65.288	0.500	0.693	20	544.017	0.0	

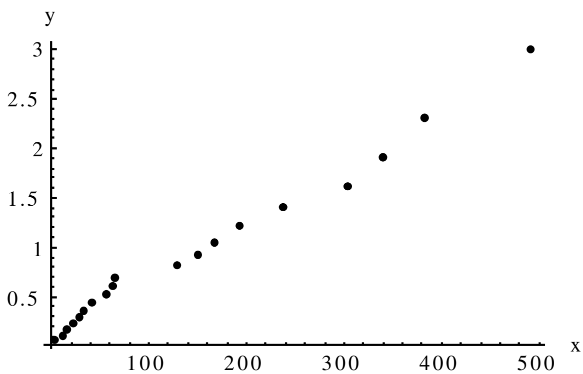


Figure 8.1 Example failure data plot.

Consider an example. Suppose 20 copies of a component are placed on test and their ordered failure times are those listed in Table 8.1. Suppose further that we calculate the corresponding values of the y_j, which are also shown in the table. Then the actual plot of the data is shown in Figure 8.1. As we can see, within the variation expected in sample data, the graph is approximately linear. Using the plot to identify the slope and thus the estimated value of λ is not particularly precise, but a reasonable choice of value appears to be $\hat{\lambda} = 0.0058$. If we compute the actual regression line, the algebraic expression for the slope is

$$\hat{\lambda} = \frac{n\sum_{j=1}^{n} x_j y_j - \sum_{j=1}^{n} x_j \sum_{j=1}^{n} y_j}{n\sum_{j=1}^{n} x_j^2 - \left(\sum_{j=1}^{n} x_j\right)^2} \tag{8.5}$$

This equation gives $\hat{\lambda} = 0.0055$ as the estimate. As a point of information, the example data was generated using a value of $\lambda = 0.005$.

Before we leave this example, note that we did not include an estimate of the reliability for the last of the observed failure times. This is because we the used mean order statistic-based estimation (Equation 7.4):

$$\hat{F}_T(t) = \frac{j}{n}$$

and in this case, $-\ln \hat{\bar{F}}_T(t)$ is undefined, as the reliability estimate is zero. If we had used the equation based on the median of the distribution on order statistics — Equation 7.5 — we would have been able to include the final data value in our calculations.

The application of Equation 8.2 to the Weibull distribution proceeds in a comparably direct manner. First, we obtain

$$-\ln \bar{F}_T(t) = \left(t/\theta \right)^{\beta}$$

and taking the logarithm again yields

$$\ln\left(-\ln \bar{F}_T(t)\right) = \beta \ln t - \beta \ln \theta \tag{8.6}$$

which is again the slope of a line. In this case, the intercept is nonzero. We again represent the dependent variable, which is the estimated reliability at each failure time, by y_j, and the result is a data set such as the example set shown in Table 8.2. Note that the values of the y_j listed in the table are computed using Equation 7.5:

Table 8.2 Example Weibull Failure Data

j	x_j	$\ln(x_j)$	$\hat{\hat{F}}_T(t)$	y_j	j	x_j	$\ln(x_j)$	$\hat{\hat{F}}_T(t)$	y_j
1	390.896	5.968	0.962	–3.246	14	932.309	6.838	0.450	–0.225
2	509.925	6.234	0.922	–2.517	15	957.288	6.864	0.411	–0.116
3	540.671	6.293	0.883	–2.085	16	984.191	6.892	0.371	–0.009
4	594.520	6.388	0.844	–1.772	17	1003.160	6.911	0.332	0.098
5	621.604	6.432	0.804	–1.524	18	1018.753	6.926	0.293	0.206
6	626.117	6.440	0.765	–1.317	19	1030.576	6.938	0.253	0.318
7	679.096	6.521	0.726	–1.137	20	1082.845	6.987	0.214	0.434
8	664.210	6.499	0.686	–0.977	21	1222.792	7.109	0.174	0.558
9	710.355	6.566	0.647	–0.831	22	1279.176	7.154	0.135	0.694
10	714.938	6.572	0.607	–0.696	23	1285.361	7.159	0.096	0.853
11	746.485	6.615	0.568	–0.570	24	1392.606	7.239	0.056	1.057
12	763.342	6.638	0.529	–0.451	25	1577.441	7.364	0.017	1.406
13	775.172	6.653	0.489	–0.336					

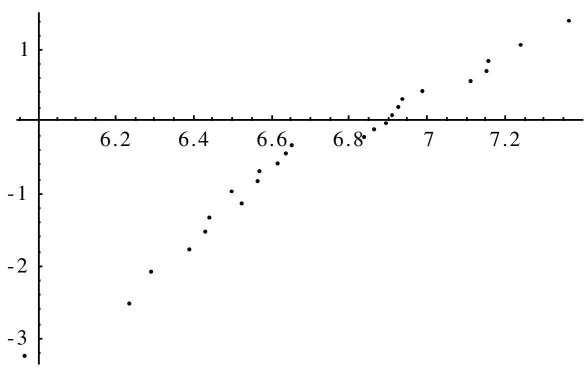


Figure 8.2 Plot of logarithms of Weibull failure data.

$$\hat{F}_T(t) = \frac{j - 0.3}{n + 0.4}$$

so all of the data may be included in our analysis. The independent variable in this case is the logarithm of the failure time as specified in Equation 8.6. The plot of the data is shown in Figure 8.2. Clearly, using the plot to obtain parameter estimates would be quite difficult for this case. Instead of trying to judge the behavior of the plot, we use the regression analysis equations to calculate our estimates.

The slope and intercept of a line fit to this type of data are given by Equations 8.7 and 8.8. For the data listed in Table 8.2, the computed value of the slope is 3.251, and that for the intercept is –22.358. Now, inverting Expression 8.6, these numerical values correspond to

$$\hat{\beta} = 3.251$$

$$\hat{\theta} = e^{-(-22.358)/\hat{\beta}} = e^{22.358/3.251} = 970.576$$

$$slope = \frac{n\sum_{j=1}^{n} y_j \ln(x_j) - \left(\sum_{j=1}^{n} y_j\right)\left(\sum_{j=1}^{n} \ln(x_j)\right)}{n\sum_{j=1}^{n} \left(\ln(x_j)\right)^2 - \left(\sum_{j=1}^{n} \ln(x_j)\right)^2} \tag{8.7}$$

$$\text{intercept} = \frac{\left(\sum_{j=1}^{n} \left(\ln(x_j)\right)^2\right)\sum_{j=1}^{n} y_j - \left(\sum_{j=1}^{n} \ln(x_j)\right)\left(\sum_{j=1}^{n} y_j \ln(x_j)\right)}{n\sum_{j=1}^{n} \left(\ln(x_j)\right)^2 - \left(\sum_{j=1}^{n} \ln(x_j)\right)^2} \tag{8.8}$$

Equations 8.7 and 8.8 are the standard linear regression forms. They have a corresponding matrix form that is actually easier to understand and use. For a general linear fit, the data pairs, say (u_j, v_j), that correspond to a model:

$$v = a + bu$$

are arranged in matrix/vector form with

$$U = \begin{bmatrix} 1 & u_1 \\ 1 & u_2 \\ \vdots & \vdots \\ 1 & u_n \end{bmatrix}$$

and

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

so that the basic model is

$$V = MU$$

with the two coordinates of the matrix M being the intercept "a" and the slope "b." The regression solution of this model for a set of data is

$$M = [U'U]^{-1}U'V \tag{8.9}$$

As noted above, for the Weibull distribution, this solution corresponds to

$$M = \begin{bmatrix} \hat{\beta} \\ \hat{\beta}\ln\hat{\theta} \end{bmatrix}$$

There are three final points related to the graphical method. The first is that the method can be applied to other distributions. However, because most other distributions used to model life length do not have a closed form representation of the cumulative distribution function, estimation of the parameters for those distributions is usually performed using methods other than the graphical ones.

Second, if a plot of Weibull data displays curvature, particularly near the ends, this is evidence of the existence of a third parameter, the minimum life parameter. This was the parameter δ in Equation 4.14. The best approach to this situation is to use a search method to identify the value of the minimum life parameter for which the regression fit is best. Consider the example of the data set given in Table 8.3. The initial linear regression plot of that data, using Equations 8.7 to 8.8, is shown in Figure 8.3. Notice the decided nonlinearity of the initial data plot and the curvature at the ends of the

Table 8.3 Example Weibull Failure Data with Minimum Life

j	x_j	$ln(x_j)$	$\hat{\hat{F}}_T(t)$	y_j	j	x_j	$ln(x_j)$	$\hat{\hat{F}}_T(t)$	y_j
1	589.614	6.379	0.966	–3.355	11	1003.984	6.912	0.475	–0.297
2	640.936	6.463	0.917	–2.442	12	1008.317	6.916	0.426	–0.160
3	670.372	6.508	0.868	–1.952	13	1030.989	6.938	0.377	–0.026
4	731.327	6.595	0.819	–1.609	14	1040.955	6.948	0.328	0.107
5	828.633	6.720	0.770	–1.340	15	1052.332	6.959	0.279	0.243
6	859.266	6.756	0.721	–1.116	16	1062.327	6.968	0.230	0.384
7	870.480	6.769	0.672	–0.921	17	1103.345	7.006	0.181	0.535
8	881.452	6.782	0.623	–0.747	18	1103.629	7.006	0.132	0.704
9	894.234	6.796	0.574	–0.587	19	1192.351	7.084	0.083	0.910
10	934.200	6.840	0.525	–0.438	20	1230.343	7.115	0.034	1.216

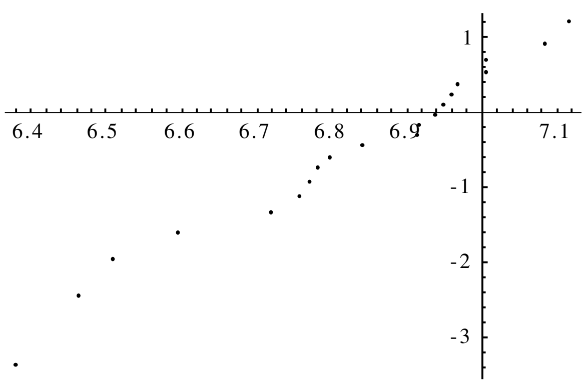


Figure 8.3 Initial plot of logarithms of Weibull failure data.

plot. By successive trials using a search strategy, we find that a value of δ = 400.0 results in the plot shown in Figure 8.4. Then applying the regression analysis to the data adjusted by δ, we obtain estimates of $\hat{\beta} = 2.85$ and $\hat{\theta} = 606.771$.

Finally, it is important to note that the graphical method (and its regression equivalent) applies directly to censored data. If, for example, a sample of n = 40 items are placed on test and the test is terminated after only 18 failures, the resulting failure times can be plotted against the corresponding estimated failure probabilities, and the parameter estimation equations are the slope and intercept expressions. As

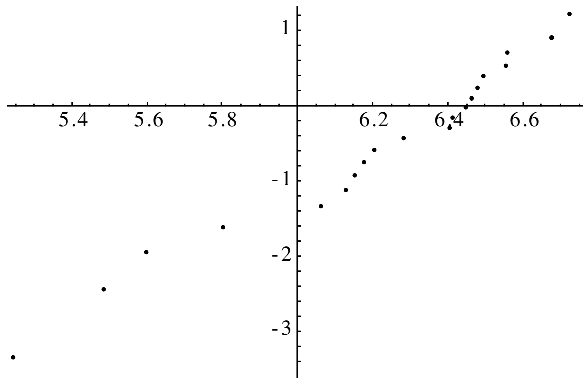


Figure 8.4 Revised plot of logarithms of Weibull failure data using δ = 400.0.

a further example, suppose only the first 12 of the data points in Table 8.1 had been observed. This means that 60% of the copies of the device on test had failed, but the total test time would then have been around 150.393 units, rather than the full 544.017 indicated in the table. This is a significant savings in test time, and the resulting parameter estimate is an identical $\hat{\lambda} = 0.0058$ to the one obtained using the full data set.

8.2 METHOD OF MOMENTS

The method of moments is actually well known and is quite intuitive. The idea is that the sample moments are equated to the algebraic expressions of the distribution moments, and the resulting equations are solved for parameter estimates. The most common realization of this idea is the use of a sample mean to estimate a population parameter such as its mean. Consider some specific cases.

For the exponential distribution, the population mean is the reciprocal of the distribution parameter:

$$E[T] = {}^{1}\!/_{\lambda} \tag{8.10}$$

As the sample mean is a reasonable estimate for the population mean, one may use $\hat{E}[t] = \bar{x}$, so

$$\hat{\lambda} = 1/\bar{x} \tag{8.11}$$

For the example data in Table 8.1, we obtain $\bar{x} = 137.23$ which yields the estimate $\hat{\lambda} = 0.0073$.

For the Weibull distribution, two equations are needed to estimate two parameters. In principle, we could use the mean and the variance expressions along with the sample mean and the sample variance. However, there is a slightly simpler approach. In general, the moments of the Weibull distribution may be determined to be

$$E[T^k] = \theta^k \Gamma\left(1 + k/\beta\right) \tag{8.12}$$

Therefore, the mean of a Weibull distribution is

$$E[T] = \theta \Gamma\left(1 + 1/\beta\right) \tag{8.13}$$

and the corresponding variance is

$$Var[T] = E[T^2] - \left(E[T]\right)^2 = \theta^2 \Gamma\left(1 + 2/\beta\right) - \left(\theta \Gamma\left(1 + 1/\beta\right)\right)^2$$

$$= \theta^2 \left(\Gamma\left(1 + 2/\beta\right) - \Gamma^2\left(1 + 1/\beta\right)\right) \tag{8.14}$$

Now, rather than equate the sample mean and sample variance to the distribution mean and variance, we take the coefficient of variation,

$$c = \frac{Var[t]}{E^2[T]} = \frac{\theta^2 \left(\Gamma\left(1 + 2/\beta\right) - \Gamma^2\left(1 + 1/\beta\right)\right)}{\theta^2 \Gamma^2\left(1 + 1/\beta\right)}$$

$$= \frac{\Gamma\left(1 + 2/\beta\right)}{\Gamma^2\left(1 + 1/\beta\right)} - 1 \tag{8.15}$$

Table 8.4
Numerical Search Values

β	c
1.0	1.0
5.0	0.0525
1.5	0.4610
3.5	0.1001
3.0	0.1321
3.3	0.1113
3.2	0.1177
3.21	0.1170
3.205	0.1173
3.208	0.11712
3.209	0.11706

which contains only one of the parameters. Given a set of failure data, we solve this expression numerically for the estimate of β, and we then use that estimate in the expression for the mean to compute an estimate for θ.

Consider an example. The failure data in Table 8.2 has a mean value of 884.153 and a sample variance of 91548 (standard deviation of 302.569). Thus, the sample value of the coefficient of variation is 0.1171. A numerical search for the value of β in Equation 8.15 that most closely matches this value yields an estimate of $\hat{\beta} = 3.208$. The sequence of functional evaluations used to find this estimate is shown in Table 8.4. Note that only 11 trials were needed to obtain the parameter estimate. Once the estimate of β is obtained, we compute the estimate of the scale parameter using the sample mean in place of the population mean in Equation 8.13:

$$\hat{\theta} = \frac{\bar{x}}{\Gamma\left(1 + \frac{1}{\hat{\beta}}\right)} = \frac{884.153}{\Gamma(1.3117)} = 987.04$$

In the case of the Normal distribution, the estimation process using the method of moments is direct, as the moments of the distribution are the mean and variance. Thus, for a sample mean of $\bar{x}$ and sample variance of s^2, the estimation equations are

$$\hat{\mu} = \bar{x} \qquad \hat{\sigma}^2 = s^2 \tag{8.16}$$

The Gamma distribution may also be analyzed using the method of moments. The approach used for the Weibull distribution is the most efficient one for the Gamma. In general, the moments of the Gamma distribution stated in Equation 4.21 are

$$E[T^k] = \frac{1}{\lambda^k}\prod_{i=0}^{k-1}\left(\beta + i\right) \tag{8.17}$$

The application of this form yields a mean value of

$$E[T] = \beta/\lambda \tag{8.18}$$

and the corresponding variance is

$$Var[T] = E[T^2] - \left(E[T]\right)^2 = \frac{\beta(\beta+1)}{\lambda^2} - \frac{\beta^2}{\lambda^2} = \beta/\lambda^2 \tag{8.19}$$

Thus, as in the case of the Weibull, the coefficient of variation,

$$c = \frac{Var[t]}{E^2[T]} = \frac{\beta/\lambda^2}{\beta^2/\lambda^2} = 1/\beta \tag{8.20}$$

is expressed in terms of a single parameter. To obtain estimates of the parameters of a Gamma distribution using sample data, we equate the reciprocal of the sample coefficient of variation to β and then use the resulting estimate along with the sample mean to estimate λ. For example, the life test data in Table 8.5 was obtained during the test of a component population for which the Gamma distribution is believed to represent well the dispersion in life lengths. As this data displays a sample mean value of 2948.75 and a sample standard

Table 8.5 Gamma Distributed Life Test Data

j	x_j	j	x_j
1	692	11	2623
2	995	12	2881
3	1239	13	2972
4	1314	14	3271
5	1530	15	3618
6	1740	16	3889
7	1949	17	4493
8	2056	18	4973
9	2199	19	6214
10	2348	20	7979

deviation of 1845.25, the application of Expression 8.20 and then Equation 8.18 yield

$$\hat{\beta} = 2.554$$

and

$$\hat{\lambda} = 0.00086$$

It should now be apparent that the method of moments is usually quite easy to apply. It has intuitive appeal. The disadvantage of the method is that it one does not expect to be able to use it with censored data. An approach for using censored data has been suggested and is discussed in the final section of this chapter.

8.3 METHOD OF MAXIMUM LIKELIHOOD

The most widely used of the parametric techniques is the method of maximum likelihood. Under this very intuitive strategy, one selects as parameter estimates those values "most likely" to have produced the observed data. Consider an example of this concept. Suppose a presumably fair coin is tossed 40 times with the result that heads appears on 16 of the trials. What estimate of the probability of heads, p, is

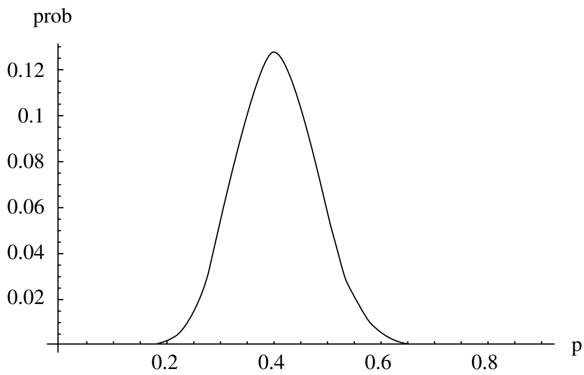


Figure 8.5 Likelihood function for a binomial sample.

appropriate? One possible answer is that since the coin is a fair coin, $\hat{p} = 0.50$. The method of maximum likelihood indicates that $\hat{p} = 0.40$. The reason for this value is that the binomial probability of observing 16 heads in 40 trials is greatest when the value of p is 0.40. That is,

$$\max_p \{b(16, 40, p\} = \max_p \left\{ \binom{40}{16} p^{16}(1-p)^{24} \right\} = 0.128$$

occurs at p = 0.40. This is illustrated in Figure 8.5 in which the binomial probability of 16 heads in 40 trials is plotted against the value of p. Note that one usually plots event probabilities against the events, but that this plot provides an alternate view of the probabilities — one focused on the parameter of the distribution.

In terms of algebraic methodology, one forms the likelihood function, which is the joint distribution of the parameter(s) and the sample outcome, and one then maximizes the function relative to the parameter value. For the binomial, the likelihood function is

$$L(x, n, p) = b(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

and to find the maximum, we take the derivative with respect to p and set it equal to zero:

$$\frac{d}{dp}L(x,\,n,\,p)=\frac{d}{dp}\binom{n}{x}p^{x}(1-p)^{n-x}$$

$$=\binom{n}{x}\left(xp^{x-1}(1-p)^{n-x}-(n-x)p^{x}(1-p)^{n-x-1}\right)=0$$

$$=x(1-p)-(n-x)p$$

$$=x-px-np+px=x-np=0$$

From this we obtain the maximum likelihood estimation equation:

$$\hat{p}=x/n$$

For our example, this means that $\hat{p}=0.40$ is the maximum likelihood estimate for the probability of heads.

The application of the principle of maximum likelihood to life test data is direct. We form the likelihood function as the joint distribution of the observed failure data and the unknown distribution parameters, and we maximize that function with respect to the parameters. That is, we form the likelihood function:

$$L(\underline{x},\,\underline{\theta})=f_T(\underline{x},\,\underline{\theta})=f_T(x_1,\,x_2,\,...,\,x_n,\,\underline{\theta}) \tag{8.21}$$

We next note that the fact that the individual failure times are mutually independent, so the joint distribution may be expressed as the product of the marginal distributions.

$$f_T(x_1,\,x_2,\,...,\,x_n,\,\underline{\theta})=\prod_{j=1}^{n}f_T(x_j,\,\underline{\theta}) \tag{8.22}$$

Then we observe further that the values of the parameters that maximize the likelihood function are the same as the ones that maximize the logarithm of the likelihood function, so in those cases in which this equivalence is useful, we can exploit it.

Consider the exponential distribution for which there is only one parameter, so $\underline{\theta} = \{\lambda\}$ and

$$f_T(x_j, \underline{\theta}) = f_T(x_j, \lambda) = \lambda e^{\lambda x_j}$$

For the exponential distribution, the likelihood function is

$$L(\underline{x}, \lambda) = \prod_{j=1}^{n} \lambda e^{-\lambda x_j} = \lambda^n e^{-\lambda \sum_{j=1}^{n} x_j} \tag{8.23}$$

and the corresponding logarithm of the likelihood function is

$$\ln\left(L(\underline{x}, \lambda)\right) = n \ln \lambda - \lambda \sum_{j=1}^{n} x_j \tag{8.24}$$

To maximize this function we take the derivative and set it equal to zero:

$$\frac{d}{d\lambda} \ln\left(L(\underline{x}, \lambda)\right) = \frac{n}{\lambda} - \sum_{j=1}^{n} x_j = 0$$

and solving for the parameter estimate yields

$$\hat{\lambda} = 1 \Big/ \frac{1}{n} \sum_{j=1}^{n} x_j = 1 / \underline{x} \tag{8.25}$$

A check of the second derivative condition indicates that the second derivative is clearly negative, so the solution we have found is a maximum. In this particular case, but not in general, the maximum likelihood and method of moments estimates are the same. As previously noted, the example data in Table 8.1 displays a mean value of $\underline{x} = 137.23$, so our estimate of the parameter is $\hat{\lambda} = 0.0073$.

In the case of the Weibull distribution, the concept is the same but there are two parameters. Thus, we form the likelihood function:

$$\begin{aligned} L(\underline{x}, \beta, \theta) &= \prod_{j=1}^{n} f_T(\underline{x}, \beta, \theta) = \prod_{j=1}^{n} \frac{\beta x_j^{\beta-1}}{\theta^{\beta}} e^{-\left(\frac{x_j}{\theta}\right)^{\beta}} \\ &= \frac{\beta^n}{\theta^{n\beta}} \left(\prod_{j=1}^{n} x_j^{\beta-1} \right) e^{-\sum_{j=1}^{n}\left(\frac{x_j}{\theta}\right)^{\beta}} \end{aligned} \tag{8.26}$$

Taking the logarithm of Equation 8.26,

$$\ln\left(L(\underline{x}, \beta, \theta)\right) = n\ln\beta - n\beta\ln\theta + (\beta-1)\sum_{j=1}^{n}\ln x_j - \frac{1}{\theta^{\beta}}\sum_{j=1}^{n} x_j^{\beta} \tag{8.27}$$

and to maximize this function, we set the partial derivatives equal to zero. We solve the resulting expressions for our parameter estimates. First,

$$\frac{\partial}{\partial\theta}\ln\left(L(\underline{x}, \beta, \theta)\right) = -\frac{n\beta}{\theta} + \frac{\beta}{\theta^{\beta+1}}\sum_{j=1}^{n} x_j^{\beta} = 0$$

$$\hat{\theta}^{\hat{\beta}} = \frac{1}{n}\sum_{j=1}^{n} x_j^{\hat{\beta}}$$

$$\hat{\theta} = \left(\frac{1}{n}\sum_{j=1}^{n} x_j^{\hat{\beta}} \right)^{1/\hat{\beta}} \tag{8.28}$$

and then

$$\begin{aligned} \frac{\partial}{\partial\beta}\ln\left(L(\underline{x}, \beta, \theta)\right) &= \frac{n}{\beta} - n\ln\theta + \sum_{j=1}^{n}\ln x_j \\ &\quad + \frac{\ln\theta}{\theta^{\beta}}\sum_{j=1}^{n} x_j^{\beta} - \frac{1}{\theta^{\beta}}\sum_{j=1}^{n} x_j^{\beta}\ln x_j = 0 \end{aligned}$$

Substituting the expression obtained for θ in this equation yields

$$\frac{1}{\hat{\beta}}+\frac{1}{n}\sum_{j=1}^{n}\ln x_j-\frac{\sum_{j=1}^{n}x_j^{\hat{\beta}}\ln x_j}{\sum_{j=1}^{n}x_j^{\hat{\beta}}}=0 \tag{8.29}$$

This final expression must be solved numerically for the value of the estimate for β, and that value is then used to compute the estimated value of θ. The second derivative conditions indicate that the solutions obtained above correspond to a maximum of the logarithm of the likelihood function. As an example, consider again the data of Table 8.2. Using a numerical search strategy, the values of the estimate for β that are evaluated are shown in Table 8.6. Observe that only 11 values were needed to obtain substantial precision in the estimate. Then, using $\hat{\beta}=3.212$ in Equation 8.28, we obtain $\hat{\theta}=988.257$. As will be shown shortly, the method of maximum likelihood can be applied to the exponential and Weibull distributions even when the data are censored. The method can be applied to the Gamma distribution, but only for complete data sets.

Table 8.6 Trial Values of the Parameter Estimate

$\hat{\beta}$	*l.h.s.*	$\hat{\beta}$	*l.h.s.*
2.0	0.288	3.20	0.002
4.0	−0.113	3.22	−0.001
3.0	0.038	3.21	0.0004
3.5	−0.046	3.215	−0.0005
3.25	−0.006	3.212	0
3.15	0.011		

Recall that, in general, the Gamma distribution does not have a closed form expression for the distribution function, but the density function is

$$f_T(t) = \frac{\lambda^\beta}{\Gamma(\beta)} t^{\beta-1} e^{-\lambda t} \tag{8.30}$$

Using this form in the likelihood function yields

$$\begin{aligned} L(\underline{x}, \beta, \lambda) &= \prod_{j=1}^{n} f_T(\underline{x}, \beta, \lambda) = \prod_{j=1}^{n} \frac{\lambda^\beta x_j^{\beta-1}}{\Gamma(\beta)} e^{-\lambda x_j} \\ &= \frac{\lambda^{n\beta}}{\left(\Gamma(\beta)\right)^n} \left(\prod_{j=1}^{n} x_j^{\beta-1} \right) e^{-\lambda \sum_{j=1}^{n} x_j} \end{aligned} \tag{8.31}$$

Taking the logarithm,

$$\begin{aligned} \ln\left(L(\underline{x}, \beta, \lambda)\right) &= n\beta \ln\lambda - n \ln\Gamma(\beta) \\ &\quad + (\beta - 1)\sum_{j=1}^{n} \ln x_j - \lambda \sum_{j=1}^{n} x_j \end{aligned} \tag{8.32}$$

and we proceed in the same manner as we did with the Weibull. We take the partial derivatives:

$$\frac{\partial}{\partial\lambda} \ln\left(L(\underline{x}, \beta, \lambda)\right) = \frac{n\beta}{\lambda} \ln - \sum_{j=1}^{n} x_j = 0$$

$$\hat{\lambda} = \beta \Big/ \frac{1}{n}\sum_{j=1}^{n} x_j = \beta / \bar{x} \tag{8.33}$$

$$\frac{\partial}{\partial\beta} \ln\left(L(\underline{x}, \beta, \lambda)\right) = n \ln\lambda - n \frac{\Gamma(\beta)}{\psi(\beta)} + \sum_{j=1}^{n} \ln x_j = 0$$

$$\ln\hat{\beta} - \ln\left(\frac{1}{n} \sum_{j=1}^{n} x_j \right) - \psi(\hat{\beta}) + \frac{1}{n} \sum_{j=1}^{n} \ln x_j = 0 \tag{8.34}$$

where the Psi function is the derivative of the logarithm of the Gamma function:

$$\psi(\beta) = \frac{d}{d\beta} \ln \Gamma(\beta) = \frac{\Gamma'(\beta)}{\Gamma(\beta)}$$

This function is relatively well behaved. There are both tables of the function and numerical strategies for computing it. One such algorithm is included in Appendix A.

The final form of Expression 8.34 was obtained by substituting Expression 8.33 into the partial derivative with respect to β. Once again, we solve numerically for the estimate for β and then use that value to compute λ using Expression 8.33. For the example data of Table 8.5, only six function evaluations are needed to obtain $\hat{\beta} = 2.97$, and using this value, we obtain $\hat{\lambda} = 0.00101$. The sucessive values of $\hat{\beta}$ obtained during the search are shown in Table 8.7. Note how rapid the convergence to the estimate is.

Table 8.7 Trial Values of the Parameter Estimate

$\hat{\beta}$	*l.h.s.*	$\hat{\beta}$	*l.h.s.*
2.0	0.093	2.9	0.005
4.0	–0.048	2.98	–0.006
3.0	–0.002	2.97	0

One final case that is quite intuitive is the Normal distribution. Taking the density function for the Normal as stated in Chapter 4, the likelihood function is

$$L(\underline{x}, \mu, \sigma^2) = \prod_{j=1}^{n} f_T(\underline{x}, \mu, \sigma^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_j-\mu)^2/2\sigma^2}$$
$$= \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j-\mu)^2} \quad (8.35)$$

and taking the logarithm yields

$$\ln L(\underline{x}, \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(x_j - \mu\right)^2 \quad (8.36)$$

The partial derivatives are

$$\frac{\partial}{\partial\sigma}\ln L(\underline{x}, \mu, \sigma^2) = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}\left(x_j - \mu\right)^2 = 0$$

$$\frac{\partial}{\partial\mu}\ln L(\underline{x}, \mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{j=1}^{n}\left(x_j - \mu\right) = 0$$

Solving the second of these equations first yields

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} x_j = \bar{x} \quad (8.37)$$

and then the first yields

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(x_j - \hat{\mu}\right)^2} \quad (8.38)$$

These expressions should seem familiar.

Returning to the exponential and Weibull distributions, the fact that these distributions have algebraic forms for their cumulative distribution functions and reliability functions permits their analysis under test censoring. Suppose a sample of n copies of a component is placed on test, and the test is terminated before all the items have failed. Suppose that only r failure times have been observed. In this case, there are r failure times $x_1, x_2, \ldots, x_r$, and $n - r$ components have survived over x_r time units. Thus, the likelihood function is

$$L(\underline{x}, \underline{\theta}) = \left(\prod_{j=1}^{r} f_T(x_j, \underline{\theta})\right)\left(\bar{F}_T(x_r)\right)^{n-r} \quad (8.39)$$

Since the exponential distribution is subsumed by the Weibull, consider the application of this form to the Weibull.

$$L(\underline{x}, \beta, \theta) = \left(\prod_{j=1}^{r} \frac{\beta x_j^{\beta-1}}{\theta^{\beta}} e^{-\left(\frac{x_j}{\theta}\right)^{\beta}} \right) \left(e^{-\left(\frac{x_r}{\theta}\right)^{\beta}} \right)^{n-r}$$
$$= \frac{\beta^r}{\theta^{r\beta}} \left(\prod_{j=1}^{r} x_j^{\beta-1} \right) e^{-\sum_{j=1}^{r}\left(\frac{x_j}{\theta}\right)^{\beta} -(n-r)\left(\frac{x_r}{\theta}\right)^{\beta}} \tag{8.40}$$

Taking the logarithm,

$$\ln\left(L(\underline{x}, \beta, \theta)\right) = r \ln\beta - r\beta \ln\theta + (\beta-1)\sum_{j=1}^{r} \ln x_j - \sum_{j=1}^{r}\left(\frac{x_j}{\theta}\right)^{\beta}$$
$$-\left((n-r)\left(\frac{x_r}{\theta}\right)^{\beta} \right) \tag{8.41}$$

The partial derivatives are

$$\frac{\partial}{\partial\theta}\ln\left(L(\underline{x}, \beta, \theta)\right) = -\frac{r\beta}{\theta} + \frac{\beta}{\theta^{\beta+1}}\left(\sum_{j=1}^{r} x_j^{\beta} + (n-r)x_r^{\beta} \right)$$

$$\frac{\partial}{\partial\beta}\ln\left(L(\underline{x}, \beta, \theta)\right) = \frac{r}{\beta} - r\ln\theta + \sum_{j=1}^{r}\ln x_j + \frac{\ln\theta}{\theta^{\beta}}\left(\sum_{j=1}^{r} x_j^{\beta} + (n-r)x_r^{\beta} \right)$$
$$-\frac{1}{\theta^{\beta}}\left(\sum_{j=1}^{r} x_j^{\beta}\ln x_j + (n-r)x_r^{\beta}\ln x_r \right)$$

Solving the first of the equations, we obtain

$$\hat{\theta} = \left(\frac{1}{r}\left(\sum_{j=1}^{r} x_j^{\hat{\beta}} + (n-r)x_r^{\hat{\beta}} \right) \right)^{1/\hat{\beta}} \tag{8.42}$$

and substituting this expression into the second of the partial derivative equations, we have

$$\frac{1}{\hat{\beta}}+\frac{1}{r}\sum_{j=1}^{r}\ln x_j-\frac{\left(\sum_{j=1}^{r}x_j^{\hat{\beta}}\ln x_j+(n-r)x_r^{\hat{\beta}}\ln x_r\right)}{\left(\sum_{j=1}^{r}x_j^{\hat{\beta}}+(n-r)x_r^{\hat{\beta}}\right)}=0 \quad (8.43)$$

Once again, we solve Expression 8.43 numerically for $\hat{\beta}$ and then compute $\hat{\theta}$ using Expression 8.42. Note that the form of Expressions 8.42 and 8.43 are consistent with those obtained for complete data sets. In fact, if we take $r = n$, these equations reduce to those obtained for complete data sets.

As an example, suppose only the first $r = 15$ of the failure times of Table 8.2 had been observed. Using the equations developed above, we obtain the parameter estimates $\hat{\beta}=3.513$ and $\hat{\theta}=965.571$.

To complete this discussion, it is appropriate to note that the method of maximum likelihood is very appealing because of its intuitive foundation. It is widely used and the fact that it can sometimes be applied to censored data is an additional positive feature of the method. The main disadvantage of the method is that the estimators obtained are not always unbiased. In particular, the estimators for the Weibull distribution are not unbiased. Even for the Normal distribution, the estimator for the mean is unbiased, but the one for the standard deviation is not.

8.4 SPECIAL TOPICS

The statistical methods for parameter estimation that have been presented in this chapter are generally adequate for nearly all test situations. However, there are three additional topics that are worthy of our attention. The first of these is the extension of the method of moments to right-censored data. As indicated above, the method of moments has historically been applied only to complete data sets. Nachlas and Kumar [43] suggest a heuristic strategy for implementing the method of moments using data sets that are censored because of test termination. The construction of the method is based on the substitution of the expected remaining life lengths of the unfailed test specimens for their actual failure times.

The second topic is the use of step-stress testing in which a sequence of progressively more severe environmental stresses are used to accelerate component aging. As is shown, the statistical analysis is essentially the same as for usual life testing. The third topic is the estimation of reliability when we have grouped data. In that case, we observe the number of surviving items at points in time rather than continuously, and we allow for the removal of some unfailed test specimens from the test at observation points. When testing proceeds in this manner, we can only obtain nonparametric reliability estimates.

8.4.1 Method of Moments with Censored Data

In the case of the exponential distribution, a test for which the first r failure times are observed and the remaining $n - r$ are not, the expression for the "surrogate" sample mean is

$$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)\big(x_r + E[Y]\big)\right) \tag{8.44}$$

where Y is the remaining life of the components that have not yet failed. The general expression for the expected value of Y has been shown by Cox [44] to be

$$E[Y] = \frac{\mu^2 + \sigma^2}{2\mu} \tag{8.45}$$

where μ is the mean of the life distribution, and σ^2 is the variance of the life distribution. For the exponential distribution, this expected value is $1/\lambda$, so Equation 8.44 becomes

$$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)\left(x_r + \frac{1}{\hat{\lambda}}\right)\right)$$

and we equate this to the population mean to solve for the estimate of the distribution parameter.

$$\frac{1}{\hat{\lambda}} = \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)\left(x_r + \frac{1}{\hat{\lambda}}\right)\right)$$

This yields the reasonably intuitive expression

$$\hat{\lambda} = \left(\frac{1}{r}\left(\sum_{j=1}^{r} x_j + (n-r)x_r\right)\right)^{-1} \tag{8.46}$$

The equation is intuitive because it represents the sample mean as the average of the observed failure times plus an additional term to compensate for the unobserved life lengths. For $r = 12$, the data in Table 8.1 yield the estimate $\hat{\lambda} = 0.0065$.

In the case of the Weibull distribution, Cox's result implies

$$E[Y] = \frac{\theta\Gamma\left(1 + 2/\beta\right)}{2\Gamma\left(1 + 1/\beta\right)}$$

and substituting this expression into Equation 8.44 yields

$$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)\left(x_r + \frac{\theta\Gamma\left(1 + 2/\hat{\beta}\right)}{2\Gamma\left(1 + 1/\hat{\beta}\right)}\right)\right)$$

$$= \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)x_r\right) + \frac{n-r}{n}\left(\frac{\theta\Gamma\left(1 + 2/\hat{\beta}\right)}{2\Gamma\left(1 + 1/\hat{\beta}\right)}\right)$$

The corresponding representation of the sample variance is most conveniently stated as

$$s^2 = \frac{1}{n-1}\left(\sum_{j=1}^{r} x_j^2 + (n-r)\left(x_r + \frac{\theta\Gamma\left(1+\frac{2}{\hat{\beta}}\right)}{2\Gamma\left(1+\frac{1}{\hat{\beta}}\right)}\right)^2 - n\bar{x}^2\right)$$

To obtain parameter estimates, the expression for $\bar{x}$ is equated to the distribution mean, and the resulting expression is solved for θ. This yields

$$\hat{\theta} = \frac{2n\left(\sum_{j=1}^{r} x_j + (n-r)x_r\right)\Gamma(1+1/\hat{\beta})}{2n\Gamma^2(1+1/\hat{\beta}) - (n-r)\Gamma(1+2/\hat{\beta})}$$

This expression is used in place of θ in the expressions for both the sample mean and the sample variance, which are included in the coefficient of variation as in the case of uncensored data. A numerical search yields an estimate for β, which is then used in the above equation to compute an estimate for θ. As an example, the first r = 15 data values from Table 8.2 have been used to obtain parameter estimates of $\hat{\beta} = 2.391$ and $\hat{\theta} = 1175.1$.

In the case of the Gamma distribution, Cox's result implies

$$E[Y] = \frac{\beta+1}{2\lambda}$$

and substituting this expression into Equation 8.44 yields

$$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)\left(x_r + \frac{\beta+1}{2\lambda}\right)\right)$$

$$= \frac{1}{n}\left(\sum_{j=1}^{r} x_j + (n-r)x_r\right) + \frac{n-r}{n}\left(\frac{\beta+1}{2\lambda}\right)$$

The corresponding representation of the sample variance is most conveniently stated as

$$s^2 = \frac{1}{n-1}\left(\sum_{j=1}^{r} x_j^2 + (n-r)\left(x_r + \frac{\beta+1}{2\lambda}\right)^2 - n\bar{x}^2\right)$$

To obtain parameter estimates, the expression for $\bar{x}$ is equated to the distribution mean, β/λ, and the resulting expression is solved for λ. This yields

$$\hat{\lambda} = \frac{2n\hat{\beta} - (n-r)(\hat{\beta}+1)}{2\left(\sum_{j=1}^{r} x_j + (n-r)x_r\right)}$$

This expression is used in place of λ in the expressions for both the sample mean and the sample variance, which are included in the coefficient of variation as in the case of uncensored data. A numerical search yields an estimate for β, which is then used in the above equation to compute an estimate for λ. As an example, the first r = 12 data values from Table 8.5 have been used to obtain parameter estimates of $\hat{\beta} = 3.540$ and $\hat{\lambda} = 0.00118$.

As should now be evident, it is possible to use right-censored data to construct estimators for distribution parameters using the method of moments. The process is fairly involved, but it does yield reasonable estimator values.

8.4.2 Data Analysis under Step-Stress Testing

Given the efficiencies associated with accelerated life testing, the idea of using a sequence of increasingly severe stress levels has been considered worthwhile for some components. The implementation of the idea is to test at each stress level for a set time interval so that the test may be characterized by the sequence $\{(s_1, \tau_1), (s_2, \tau_2), \ldots, (s_m, \tau_m)\}$ in which the s_i are the stress levels and the τ_i are the change times [45, 46].

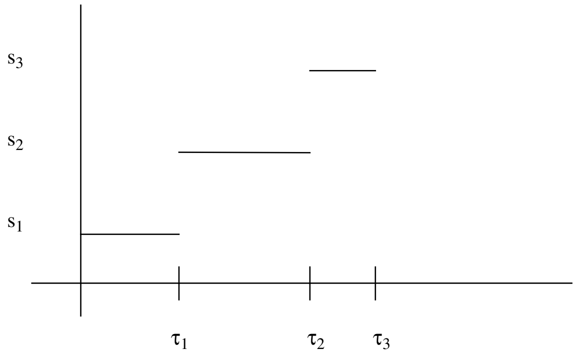


Figure 8.6 Representation of a three-level step-stress regimen.

For a sequence of three stress levels, the progress of the test is illustrated in Figure 8.6. Note that the interval over which each stress is applied is $\tau_i - \tau_{i-1}$.

There is a significant risk in the use of the step stress approach. The risk is that an excessive stress level (an overstress) will be used, and that this will alter the nature of the failure process and thus change the life distribution. Usually, the motivation for using a step-stress test is to provoke failures very rapidly in order to study a new design for which the failure behavior is not yet understood. In this case, the risk of overstress and erroneous inference is substantial. Thus, the use of a step-stress regimen must be undertaken carefully.

Assuming, as we did for a single stress, that the life distribution under all stresses is shape invariant and that we avoid the application of an excessive stress level, the data from a step-stress test permits two useful analyses. First, we can compute an acceleration factor for each stress level and map the observed failure times to the life distribution under normal operating conditions. Second, we can use the test data to estimate physical parameters such as activation energy.

If the acceleration model is already known or assumed, and a set of test data is obtained in the form of observed failure times $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$, then the corresponding equivalent data set, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, is defined by

Table 8.8 Example Data for a Step Stress Test

j	t_j	x_j	j	t_j	x_j
1	1.316	14.056	14	52.155	689.435
2	3.411	36.427	15	56.620	785.784
3	6.031	64.396	16	61.629	893.898
4	9.122	97.403	17	67.305	1016.375
5	12.675	135.346	18	73.811	1156.797
6	16.702	178.343	19	81.387	1320.296
7	21.228	226.667	20	90.390	1514.574
8	26.290	280.722	21	101.389	1751.944
9	31.940	341.051	22	143.300	2723.871
10	38.242	408.348	23	160.333	3439.661
11	41.236	453.784	24	191.578	4752.684
12	44.524	524.749	25	No failure	
13	48.146	602.917			

$$x_j = \begin{cases} a_1 t_j & 0 < t_j \le \tau_1 \\ a_1\tau_1 + a_2(t_j - \tau_1) & \tau_1 < t_j \le \tau_2 \\ a_1\tau_1 + a_2(\tau_2 - \tau_1) + a_3(t_j - \tau_2) & \tau_2 < t_j \le \tau_3 \\ \vdots & \vdots \\ a_1\tau_1 + a_2(\tau_2 - \tau_1) + \ldots + a_m(t_j - \tau_{m-1}) & \tau_{m-1} < t_j \le \tau_m \end{cases} \quad (8.47)$$

Then, this data set may be analyzed using any of the parametric (or nonparametric) methods described previously.

Consider an example. Suppose $n = 25$ copies of a device for which life length is temperature dependent are subjected to a step-stress test with the regimen {(85°C, 40 h), (95°C, 100 h), (105°C, 60 h)}. Suppose further that the device has an activation energy of 0.80 eV/°K, and that the failure times recorded during the test are those shown in Table 8.8. The corresponding equivalent failure times are also shown in the table. Subjecting the equivalent failure times to the method of moments estimation procedure for a Weibull distribution yields the parameter estimates of $\hat{\beta} = 0.7782$ and $\hat{\theta} = 986.198$.

An alternate use of data generated in an accelerated life test under a step-stress regimen is the estimation of the parameters of the physical model of failure. Specifically, we can compute jointly the parameter estimates and an estimate for the activation energy. We can use the data in Table 8.8 to illustrate the computations. Assume the activation energy is not known and that we have recorded only the "raw failure times," t_j.

Using the data from the first 40 h of testing, we obtain graphical estimates of the distribution parameters at a temperature of 85°C of $\hat{\beta} = 0.834$ and $\hat{\theta} = 87.632$. Now, the failure times after 40 hours are obtained at a higher temperature, so aging is accelerated relative to the temperature of 85°C. Assuming shape parameter invariance of the life distribution, we can compute the times at which the reliability should take values corresponding to say 12, 9, or 6 (or any number between 15 and 4) survivors from a set of 25 devices on test. With the above parameter estimates, we obtain

$$\bar{F}_{T_1}(t_{13}^e) = \frac{12.7}{25.4} = 0.5 = e^{-\left(t_{13}^e / 87.632\right)^{0.834}}$$

so without acceleration, $t_{13}^e = 56.469$

$$\bar{F}_{T_1}(t_{16}^e) = \frac{9.7}{25.4} = 0.382 = e^{-\left(t_{16}^e / 87.632\right)^{0.834}}$$

so without acceleration, $t_{16}^e = 83.719$

$$\bar{F}_{T_1}(t_{19}^e) = \frac{6.7}{25.4} = 0.264 = e^{-\left(t_{19}^e / 87.632\right)^{0.834}}$$

so without acceleration, $t_{19}^e = 123.651$

where t_j^e is our "extrapolated" value for t_j. Now, we only need one of these, but note that for all three

$$a(t_j - \tau) = t_j^e - \tau \tag{8.48}$$

where τ = 40 hours is the test time at the initial test temperature. This expression follows directly from the relation that $\bar{F}_a(t) = \bar{F}_T(at)$. In any case, the extrapolated and observed failure times yield

$$a(t_{13} - \tau) = a(48.146 - 40.0) = t_{13}^e - \tau = 56.469 - 40.0$$

so $$a = 2.021$$

$$a(t_{16} - \tau) = a(61.629 - 40.0) = t_{13}^e - \tau = 83.719 - 40.0$$

so $$a = 2.021$$

$$a(t_{19} - \tau) = a(48.146 - 40.0) = t_{19}^e - \tau = 123.651 - 40.0$$

so $$a = 2.021$$

Thus, the acceleration associated with increasing the temperature from 85°C to 95°C is 2.021. Using the Arrhenius acceleration equation,

$$a = 2.021 = e^{\frac{E_a}{K}\left(\frac{1}{T_0} - \frac{1}{T_a}\right)}$$

$$E_a = 0.801$$

Naturally, once the activation energy is estimated, the characteristic life at normal use temperature may be computed directly. For a normal use temperature of 55°C, the acceleration at 85°C is 10.678, so $\hat{\theta} = 935.734$.

8.4.3 Data Analysis with Group Censoring

There are situations in which a sample of copies of a device are put on test, and the test specimens cannot or simply are not monitored continuously. Instead, the test specimens are examined periodically to see how many are still surviving, and the data recorded is the number of failed devices. In the most general case, the observation times are also considered to provide an opportunity to remove some specimens from the test. Also, while it is most common to use equal length intervals, the general construction can be based on intervals of unequal lengths.

Suppose a total of n copies of a device are placed on test (or in operation) and that the status of the specimens is inspected at each of the times $(\tau_1, \tau_2, \ldots, \tau_k)$. Let x_j represent the number of specimens found to be failed at inspection time τ_j, and assume that, at that time, we remove m_j of the specimens from the test. Thus, the number of specimens that are tested during any interval, say the j^{th} interval, is

$$n_j = n - \sum_{i=1}^{j-1} (x_i + m_i) \tag{8.49}$$

and the conditional probability of failure during the j^{th} interval may be estimated by

$$\hat{p}_j = x_j \Big/ n_j = x_j \Big/ \left(n - \sum_{i=1}^{j-1} (x_i + m_i) \right) \tag{8.50}$$

The survivor function at any inspection time τ_j may be estimated as

$$\hat{\bar{F}}_T(\tau_j) = \prod_{i=1}^{j} (1 - \hat{p}_i) = \prod_{i=1}^{j} \left(1 - x_i \Big/ \left(n - \sum_{l=1}^{i-1} (x_l + m_l) \right) \right) \tag{8.51}$$

The variance of this survivor function estimate may also be estimated using Greenwood's formula [36]:

$$\hat{Var}\left(\hat{\bar{F}}_T(\tau_j)\right) = \left(\hat{\bar{F}}_T(\tau_j)\right)^2 \sum_{i=1}^{j} \frac{\hat{p}_i}{n_i(1 - \hat{p}_i)} \tag{8.52}$$

Once we have estimates of the reliability and the variance of the estimates, the Normal distribution confidence intervals described in Section 7.3 may be applied directly.

Consider an example. Meeker and Escobar [47] present heat exchanger tube failure data in which $n = 300$, $\boldsymbol{m} = (99, 95)$ and $\boldsymbol{x} = (4, 5, 2)$. The three intervals are each 1 year long. Using this data, $n_1 = 300$, $n_2 = 197$, and $n_3 = 97$. Therefore,

$$\hat{p}_1 = \frac{4}{300} = 0.0133$$

$$\hat{p}_2 = \frac{5}{197} = 0.0254$$

$$\hat{p}_3 = \frac{2}{97} = 0.0253$$

Then,

$$\hat{\bar{F}}_T(\tau_1) = 1 - \hat{p}_1 = 0.9867$$

and $$\hat{Var}\left(\hat{\bar{F}}_T(\tau_1)\right) = 4.385 \times 10^{-5}$$

$$\hat{\bar{F}}_T(\tau_2) = (1 - \hat{p}_1)(1 - \hat{p}_2) = 0.9616$$

and $$\hat{Var}\left(\hat{\bar{F}}_T(\tau_2)\right) = 1.578 \times 10^{-4}$$

$$\hat{\bar{F}}_T(\tau_3) = (1 - \hat{p}_1)(1 - \hat{p}_2)(1 - \hat{p}_3) = 0.9418$$

and $$\hat{Var}\left(\hat{\bar{F}}_T(\tau_3)\right) = 3.438 \times 10^{-4}$$

Finally, 95% confidence intervals for the survivor function values are

$$0.9737 \le \bar{F}_T(\tau_1) \le 0.9996$$

$$0.9370 \le \bar{F}_T(\tau_2) \le 0.9862$$

$$0.9054 \le \bar{F}_T(\tau_3) \le 0.9781$$

8.5 EXERCISES

1. Assume the following data has been generated in the life test of a component that is believed to have exponential life distribution. Use the complete data set to obtain a graphical estimate of the distribution parameter.

30.950	343.959	118.522	117.887	21.685
68.513	193.503	42.602	64.028	60.502
319.443	200.655	258.156	205.073	247.614
135.314	282.421	5.153	227.674	103.803
211.208	426.444	240.096	15.180	536.120
425.731	433.697	79.7491	258.614	26.505
541.114	8.043	75.666	272.753	50.566
247.343	379.959	220.658	415.477	211.587
123.526	324.697	54.672	67.085	77.181
61.702	1011.308	104.253	547.470	237.281

2. Use the data of Problem 1 to compute the method of moments and the maximum likelihood estimates of the distribution parameter.
3. Assume that only the earliest 20 data values of the data set of Problem 1 are available. Use those values to compute the estimates of the distribution parameter using each of (a) the graphical method, (b) the maximum likelihood method, and (c) the method of moments.
4. Assume the following data has been generated in the life test of a component that is believed to have exponential life distribution. Use the complete data set to obtain a graphical estimate of the distribution parameter.

158.033	9.964	247.922	324.953	984.249
579.983	268.967	211.775	252.635	185.474
142.382	524.376	455.421	268.403	9.362
73.521	6.956	145.474	9.527	58.538
799.290	274.426	1500.586	229.539	45.400
2609.448	175.002	50.447	252.747	847.613
711.720	491.442	982.752	2894.425	61.204
505.976	416.891	138.747	12.911	392.515
12.908	497.851	1280.381	213.168	648.615
302.120	780.516	109.058	875.372	197.162

5. Use the data of Problem 1 to compute the method of moments and the maximum likelihood estimates of the distribution parameter.
6. Assume that only the earliest 20 data values of the data set of Problem 2 are available. Use those values to compute the estimates of the distribution parameter using each of (a) the graphical method, (b) the maximum likelihood method, and (c) the method of moments.
7. Assume the following data has been generated in the life test of a component that is believed to have Weibull life distribution. Use the complete data set to obtain graphical estimates of the distribution parameters.

176.225	1058.727	918.823	29.835	203.809
252.424	695.579	68.348	1262.860	309.879
333.961	2.970	76.218	191.423	3094.748
141.948	75.239	1048.132	804.623	55.853
1989.75	1.340	641.953	152.969	126.097
1646.635	346.605	762.882	691.363	434.761
344.135	801.645	206.062	654.951	382.691
48.831	154.884	1122.269	409.087	170.046
480.549	103.444	593.04	1615.691	1880.473
131.215	724.292	224.793	470.408	957.271

8. Use the data of Problem 7 to compute the method of moments and the maximum likelihood estimates of the distribution parameters.

9. Assume that only the earliest 20 data values of the data set of Problem 2 are available. Use those values to compute the estimates of the distribution parameter using each of (a) the graphical method, (b) the maximum likelihood method, and (c) the method of moments.
10. Assume the following data has been generated in the life test of a component that is believed to have Weibull life distribution. Use the complete data set to obtain graphical estimates of the distribution parameters.

170.998	216.222	335.464	456.731	635.655
465.023	204.064	100.790	459.482	369.012
254.582	228.195	453.539	420.944	312.489
319.789	528.971	82.843	306.064	196.092
285.048	270.256	356.053	216.330	72.393
307.343	255.021	181.568	137.704	22.150
318.541	302.217	93.694	180.638	302.257
242.783	117.524	314.594	159.855	114.434
458.005	70.280	282.015	231.442	68.381
130.900	93.151	172.954	203.094	200.899

11. Use the data of Problem 10 to compute the method of moments and the maximum likelihood estimates of the distribution parameters.
12. Assume that only the earliest 25 data values of the data set of Problem 10 are available. Use those values to compute the estimates of the distribution parameter using each of (a) the graphical method, (b) the maximum likelihood method, and (c) the method of moments.
13. Assume the following data has been generated in the life test of a component that is believed to have Gamma life distribution. Use the complete data set to obtain the maximum likelihood and the method of moments estimates of the distribution parameters.
14. Assume that only the earliest 12 data values of the data set of Problem 13 are available. Use those values to compute the estimates of the distribution param-

2765.078	5218.866	1283.310	3893.011
4976.570	3472.228	1461.590	2779.887
1982.422	2262.938	777.843	2106.880
3038.784	838.939	4147.827	4622.490
2480.552	6357.559	2281.027	4269.130

eter using each of (a) the maximum likelihood method and (b) the method of moments.

15. Assume the following data has been generated in the life test of a component that is believed to have Normal life distribution. Use the complete data set to obtain the maximum likelihood and the method of moments estimates of the distribution parameters.

404.271	399.481	377.337	372.954
421.809	407.544	407.888	416.883
391.095	403.850	443.486	408.798
402.554	411.579	414.333	394.265
419.868	380.270	400.514	372.843

16. Suppose 42 copies of a device for which temperature dependent life length is believed to be well represented by a Weibull distribution were subjected to a step-stress accelerated life test with a stress regimen of {(85°C, 40 h), (95°C, 80 h), (105°C, 40 h)} and that the data in the following table were obtained. Compute estimates for the distribution parameters at the

normal operating temperature of 85°C and for the activation energy.

j	t_j	j	t_j
1	0.3799	22	40.5757
2	0.3894	23	44.3383
3	0.5062	24	48.9265
4	1.3523	25	49.5716
5	2.9877	26	51.2090
6	3.0075	27	51.5310
7	3.1151	28	53.6824
8	3.4273	29	54.7999
9	4.3191	30	63.3631
10	6.2674	31	64.6060
11	6.6553	32	69.4093
12	8.3147	33	69.6571
13	13.8439	34	75.79811
14	15.7537	35	101.3423
15	16.2473	36	110.6855
16	17.7215	37	117.9570
17	22.3923	38	125.9469
18	23.1186	39	129.8984
19	24.5316	40	130.6095
20	26.7895	41	140.1335
21	36.8783	42	No failure

17. Test data for a prototype battery [47] were accumulated over 600 hours with an initial test sample of $n = 68$ specimens. The batteries were only inspected at 50-hour intervals so $\boldsymbol{\tau} = (50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600)$. The associated observed failures and specimen removals were $\boldsymbol{x} = (1, 0, 1, 4, 1, 1, 1, 4, 4, 2, 2)$ and $\boldsymbol{m} = (5, 6, 1, 6, 2, 1, 2, 2, 3, 1, 0)$. Compute reliability estimates for each of the inspection times and confidence intervals for the estimates at 200, 400, and 600 hours.

9

Repairable Systems I — Renewal and Instantaneous Repair

Most modern equipment is designed under the assumption that it will be maintained in some manner. Complicated equipment is usually expected to operate for substantial lengths of time, so upkeep and repair activities are assumed to be part of the device operating experience. A familiar example is the personal automobile, which is comprised of many components, is expensive, and is usually expected to operate properly for several years if serviced appropriately.

A comprehensive examination of the operation of repairable systems must necessarily include a wide variety of equipment behaviors. We consider that there are essentially two classes of maintenance — preventive and corrective — and for each of these classes, there are numerous specific policy realizations. There are also several classes of possible outcomes of maintenance activity. In addition, the maintenance activity itself is subject to model representation.

The discussion of repairable systems starts here with a consideration of corrective maintenance only. In addition, repair is assumed to be instantaneous and to yield device renewal. In the chapters that follow this one, repair times of random duration are considered, nonrenewal models are discussed, and both instantaneous preventive maintenance and preventive maintenance of random duration are treated.

A defining feature of the analyses of the present chapter is that the state of the system or component of interest is assumed to be the same as new following a maintenance activity. The assumption of good as new implies device renewal. As will be discussed subsequently, the assumption about the state of a system following maintenance is one of the principal defining features of models of the behavior of repairable systems. The foundations for both the renewal and the nonrenewal cases are treated in the present chapter.

To start our examination of models of the operation of repairable systems, consider the simple definition:

> *Defn. 9.1*: A *repairable system* is an equipment entity that is capable of being restored to an operating condition following a failure.

While simple, this definition permits us to clearly distinguish between our models of life length prior to failure and the models we now develop to represent periods of operation that may extend across several failures and multiple life lengths.

Naturally, to build our models, we start with the most basic forms and then expand their features. The logical first step in studying repairable systems is to represent the simplest sequence of operating periods. We use a renewal process for this model so we begin with a review of some concepts from renewal theory and some corresponding classifications of life distributions. We then move on to the more complicated models.

9.1 RENEWAL PROCESSES

A key feature in our study of repairable systems is that we change the way in which we view the system and in which we use probabilities. Rather than consider simple life lengths and their associated distributions, we consider a series of operating intervals, and we say that the end points of these intervals form a *point process*. Naturally, the duration probabilities for the intervals of the point processes are based on the life distributions. The simplest general stochastic model of a representative point process is that for a renewal process. We start with this one. Formally,

> *Defn. 9.2*: A *renewal process* is an indexed sequence of independent and identically distributed nonnegative random variables, say T_1, T_2, ….

Conceptually, we think in terms of a device that is operated until it fails, at which time it is immediately replaced with an identical but new device that is, in turn, operated until failure. For the moment, we consider that the replacement is accomplished instantaneously, so the duration of the replacement activity can be ignored. Because they are mutually independent and have the same distribution function on length, the sequence of operating times forms a renewal process.

The above rather terse definition of a renewal process is assumed to imply that the random variables of the sequence have a well-defined probability distribution function and that this distribution function does not change over time (or as the process continues). A distinction that is often made when the process is a sequence of device operating times is that the first device in use may be new or used. (The process may have started before we began observing it.) If the first device is new, then T_l has the same distribution as all other T_i, and the process is said to be an "ordinary" renewal process. On the other hand, if the first device is used, it may not have exactly the same life distribution as the remaining copies of the device. (It has a residual life distribution.) In this case, the renewal process is said to be "modified" or "delayed." For now, consider only ordinary renewal processes.

Note also that the definition of a renewal process is the starting point for an extensive body of study of which reliability and equipment maintenance form only a small part. It is a rich model format that has many applications.

A simple graphical representation of the point process corresponding to the operation of a sequence of identical devices is shown in Figure 9.1. Note that the duration of each of the successive intervals is denoted by T_i, while the variables S_k represent the total time elapsed until the time of the k^{th} replacement. That is

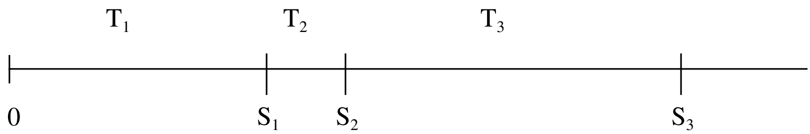


Figure 9.1 Illustration of a renewal process.

$$S_k = \sum_{i=1}^{k} T_i \tag{9.1}$$

and we say that S_k is the duration of the interval over which k copies of the device are operated to failure. As each of the T_i has the same probability distribution function, $F_T(t)$, the probability distribution on S_k is the k-fold convolution of $F_T(t)$. We represent this as

$$F_{S_k}(t) = F_T^{(k)}(t) \tag{9.2}$$

where

$$F_T^{(k)}(t) = \int_0^t F_T^{(k-1)}(t-u) f_T(u) du \tag{9.3}$$

and by convention, $F_T^{(0)}(t) = 1$ for all values of time.

The conceptual application of this model form to individual system components is straightforward. We may consider that the model represents the sequence of copies of a component that are used in a particular position (or slot) in a system structure. For example, we might consider that it represents the sequence of lightbulbs used in a lamp.

There are two natural questions that one can address using this basic model. These are (1) how long until the k^{th} failure? and (2) how many failures will occur over a fixed time interval? That is, how soon and how many? It is easy to see that the answers to these questions form the basis for planning the volume of spare parts that are purchased, the extent of investment in repair facilities and equipment, the levels of staffing that are established, and perhaps, the extent to which substitute systems are acquired.

The general expression of these questions is that we wish to examine the time until an arbitrary number of renewals have occurred and the probabilities on the number of renewals over a defined time interval. The probabilities for these two measures are related by the important fundamental relation

$$\Pr[N_t \geq k] = \Pr[S_k \leq t] \tag{9.4}$$

where N_t represents the number of renewals that occur during the interval $[0, t]$. This expression is a bit subtle and is worth pondering. An example realization is that the number of renewals during 1000 hours can equal or exceed four only if the time of the fourth renewal is equal to or earlier than t = 1000 hours.

We exploit this relationship to address the questions of how soon and how many. Note also that we are really considering a measure in the "time domain" and a measure in the "frequency domain," and that Equation 9.4 is the bridge between the two. One further point is that, once we have used Equation 9.4 to address the questions of time until the k^{th} renewal and the number of renewals over time, we can use the results of that analysis to determine:

1. The expected number of renewals during an interval
2. The identity of the renewal density
3. Higher moments of the distribution on renewals
4. The distribution of backward recurrence times
5. The distribution on forward recurrence times

The forward recurrence time is the time from an arbitrary point in time until the next event (failure), and the backward recurrence time is the time that has elapsed since the last event.

To exploit Equation 9.4, we assume that we know the distribution $F_T(t)$ on the length of the individual operating intervals. In principle, this means we also know the distribution on S_k, as we presume that we can construct the convolution of $F_T(t)$. Then, to determine the probability distribution on the number of renewals, we use

$$\Pr[N_t = n] = \Pr[N_t \geq n] - \Pr[N_t \geq n+1] = F_T^{(n)}(t) - F_T^{(n+1)}(t) \quad (9.5)$$

Observe that one realization of Equation 9.5 is

$$\Pr[N_t = 0] = F_T^{(0)}(t) - F_T^{(1)}(t) = 1 - F_T(t) = \bar{F}_T(t)$$

Now, to proceed with the identification of the specific forms of the distributions on N_t and S_k, we must specify $F_T(t)$. The most well-known construction of this type, and one of the few that is tractable at all, starts with the selection of the exponential distribution for $F_T(t)$. In this special case, the point process is called a Poisson Process.

If the lengths of the individual intervals are exponential in distribution, then the variable S_k has a Gamma distribution. This result is relatively well-known, but its construction is repeated here to remind the reader of the method of analysis. We start with the statement of the exponential distribution as

$$F_T(t) = 1 - e^{-\lambda t}$$

and we state its corresponding Laplace transform as

$$f_T^*(s) = L_{f_T}(s) = L\left(f_T(t)\right) = \int_0^\infty e^{-st} f_T(t)dt = \int_0^\infty \lambda e^{-st-\lambda t} dt = \frac{\lambda}{s+\lambda}$$

for the density function and

$$F_T^*(s) = L\left(F_T(t)\right) = \frac{\lambda}{s(s+\lambda)}$$

for the distribution function. Often the transform for the density is easier to use, so we proceed with it.

The transform for the convolution is the product of the transforms for the distributions included in the convolution. This means that the transform for the distribution on S_k is the product of k identical terms, each of which is the transform on the exponential distribution. Thus,

$$f_{S_k}^*(s) = \left(f_T^*(s)\right)^k = \left(\frac{\lambda}{s+\lambda}\right)^k$$

This is the transform for a Gamma distribution, so

$$f_{S_k}(t) = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}$$

and since k is clearly an integer, we have

$$F_{S_k}(t) = \sum_{j=k}^{\infty} \frac{(\lambda t)^j}{\Gamma(j+1)} e^{-\lambda t} \tag{9.6}$$

The corresponding realization of Equation 9.5 is

$$\Pr[N_t = n] = F_T^{(n)}(t) - F_T^{(n+1)}(t) = \sum_{j=n}^{\infty} \frac{(\lambda t)^j}{\Gamma(j+1)} e^{-\lambda t} - \sum_{j=n+1}^{\infty} \frac{(\lambda t)^j}{\Gamma(j+1)} e^{-\lambda t}$$

$$\Pr[N_t = n] = \frac{(\lambda t)^n}{\Gamma(n+1)} e^{-\lambda t} \tag{9.7}$$

In other words, if the durations of the individual operating intervals are exponentially distributed, then the time until the n^{th} renewal is Gamma distributed, and the number of renewals over a fixed time interval has a Poisson distribution.

The example of the exponential case is a bit misleading. In general, the construction of the distribution on the number of renewals is much more difficult. In fact, there are many cases in which one cannot obtain a tractable solution. Nevertheless, the review of the procedure is worthwhile, and it leads us to examine the expectation on the number of renewals that can often be obtained even if the distribution cannot be constructed.

The expected number of renewals over a fixed time interval is called the renewal function. It is the expectation for the distribution on N_t and is defined as

$$M_{F_T}(t) = E[N_t] = \sum_{n=0}^{\infty} n \Pr[N_t = n]$$
$$= \sum_{n=0}^{\infty} n\left(F_T^{(n)}(t) - F_T^{(n+1)}(t)\right) = \sum_{n=1}^{\infty} F_T^{(n)}(t) \tag{9.8}$$

The analysis of the renewal function is often accomplished by first transforming it into a recursive form:

$$M_{F_T}(t) = \sum_{n=1}^{\infty} F_T^{(n)}(t) = F_T^{(1)}(t) + \sum_{n=2}^{\infty} F_T^{(n)}(t)$$
$$= F_T(t) + \sum_{j=1}^{\infty} F_T^{(j+1)}(t)$$
$$= F_T(t) + \sum_{j=1}^{\infty} \int_0^t F_T^{(j)}(t-u) f_T(u) du$$
$$= F_T(t) + \int_0^t \sum_{j=1}^{\infty} F_T^{(j)}(t-u) f_T(u) du$$

$$M_{F_T}(t) = F_T(t) + \int_0^t M_{F_T}(t-u) f_T(u) du \tag{9.9}$$

This is the very well-known fundamental form known as the *key renewal theorem*. Very often, this form serves as the basis for the analysis of a renewal process. The main reason for its extensive use is that the associated Laplace transform yields a direct relationship between the renewal function and the underlying distribution of the process $F_T(t)$. That relationship is

$$M_{F_T}^*(s) = F_T^*(s) + L\left(\int_0^t M_{F_T}(t-u) f_T(u) du\right)$$
$$= F_T^*(s) + M_{F_T}^*(s) f_T^*(s) \tag{9.10}$$

The transform for the integral is shown in Appendix C. Then, we have

$$M^*_{F_T}(s) = \frac{F^*_T(s)}{1 - f^*_T(s)} = \frac{F^*_T(s)}{1 - s\mathrm{F}^*_T(s)} \tag{9.11}$$

and equivalently,

$$F^*_T(s) = \frac{M^*_{F_T}(s)}{1 + sM^*_{F_T}(s)} = \frac{M^*_{F_T}(s)}{1 + m^*_{F_T}(s)} \tag{9.12}$$

Of course, the utility of these results depends upon our ability to invert the transforms in any particular application.

Note that in Equation 9.12, the lower case m represents the renewal density. The renewal density is the first derivative of the renewal function. As a derivative, the renewal density necessarily represents a rate. It is the rate at which renewals occur. Thus, it is the rate at which the number of failures (renewals) increases. It is the "rate of failure." The quotation marks are intended to signal the fact that there is considerable opportunity for misunderstanding here.

Some people refer to the hazard function for a life distribution as the failure rate, while others use the term failure rate to mean the renewal density. These two entities are not the same conceptually, and except for the case of the exponential distribution, they are not the same algebraically. As explained in Chapter 4, the hazard function is the *conditional* probability of failure for members of a population given survival to any time. As shown above, the renewal density is the *unconditional* probability of another event in a sequence of events. In an effort to clearly distinguish between the two entities, the term failure rate is not used here. The hazard function is called the hazard function or hazard rate, and the renewal density is called the renewal density, the intensity function, or the failure intensity. In fact, the most appropriate label for the renewal density (aside from renewal density) is the failure intensity.

The renewal density really is the intensity with which new renewals occur. Algebraically, we can represent the derivative as

$$m_{F_T}(t) = \frac{d}{dt} M_{F_T}(t) \tag{9.13}$$

and we note that the derivative extends to both Equation 9.8 and the key renewal theorem. Hence, we have

$$m_{F_T}(t) = \sum_{n=1}^{\infty} f_T^{(n)}(t) \tag{9.14}$$

$$m_{F_T}(t) = f_T(t) + \int_0^t m_{F_T}(t-u) f_T(u)du \tag{9.15}$$

and
$$m_{F_T}^*(s) = \frac{f_T^*(s)}{1 - f_T^*(s)}$$

and
$$f_T^*(s) = \frac{m_{F_T}^*(s)}{1 + m_{F_T}^*(s)} \tag{9.16}$$

Clearly, depending upon their relative complexity, one may work with either the renewal function or the renewal density.

The single case in which the analysis is relatively straightforward is the exponential case. As indicated above, the forms for the Laplace transforms for the exponential are

$$f_T^*(s) = \frac{\lambda}{s + \lambda}$$

and
$$F_T^*(s) = \frac{\lambda}{s(s + \lambda)}$$

so

$$m_{F_T}^*(s) = \frac{\lambda}{s}$$

and
$$M_{F_T}^*(s) = \frac{\lambda}{s^2}$$

In this case, the inversion of the transform is easy and yields

$$m_{F_T}(t) = \lambda$$

and

$$M_{F_T}(t) = \lambda t \tag{9.17}$$

Note that these results imply that when the individual interval lengths are exponentially distributed, the expected number of renewals (failures) over an interval of length t is λt, and the intensity with which new failures occur is λ. Thus, for the Poisson Process, the failure intensity is constant.

For many other life distributions, the analysis of the renewal function is rather difficult. In those cases, there are some basic results that can be useful. A few of those results are

$$\begin{gathered} M_{F_T}(t) \geq \frac{t}{E_F[T]} - 1 \\ \lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{E_F[T]} \\ \lim_{t\to\infty} \frac{M_{F_T}(t)}{t} = \frac{1}{E_F[T]} \\ \lim_{t\to\infty} \Big(M_{F_T}(t+x) - M_{F_T}(t) \Big) = \frac{x}{E_F[T]} \end{gathered} \tag{9.18}$$

where the notation $E_F[T]$ is used to represent the mean of the distribution on the lengths of the individual intervals, and it is assumed that mean is finite. As an example, applying these expressions to the exponential distribution for which the expected value is $1/\lambda$ yields

$$\begin{gathered} M_{F_T}(t) \geq \lambda t - 1 \\ \lim_{t\to\infty} \frac{N_t}{t} = \lambda \\ \lim_{t\to\infty} \frac{M_{F_T}(t)}{t} = \lambda \\ \lim_{t\to\infty} \Big(M_{F_T}(t+x) - M_{F_T}(t) \Big) = \lambda x \end{gathered}$$

The corresponding results for a Weibull life distribution having shape parameter $\beta = 1.5$ and scale parameter $\theta = 2500$ hours are

$$M_{F_T}(t) \geq \frac{t}{2256.86} - 1$$

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{2256.86} = 4.43 \times 10^{-4}$$

$$\lim_{t\to\infty} \frac{M_{F_T}(t)}{t} = 4.43 \times 10^{-4}$$

$$\lim_{t\to\infty} \left(M_{F_T}(t+x) - M_{F_T}(t) \right) = \frac{x}{2256.86}$$

Note particularly the fact that Expression 9.18 applies to nearly any choice of distribution, so the above results can be very useful. The last of the four relationships is known as Blackwell's Theorem. It is well treated in Feller [48]. In effect, that last relationship states that, as time advances, a renewal process "settles down" and experiences renewals about once per $E_F[T]$, so the expected number of renewals during any interval is the length of the interval divided by the mean of the underlying distribution.

To close this discussion, note that a sequence of intervals over which a series of copies of a component are used in a machine can reasonably be represented using the renewal process model, provided each interval has the same stochastic characteristics. When one is examining components and even some modules, this is often the case. For other levels of equipment aggregation and other types of operational profiles, we will modify the renewal models later in this text. Note also that several of the most popular choices of distribution for representing operating durations yield renewal functions that are impossible or quite taxing to analyze. However, modern computing power has made these distributions much more manageable. The example of the Weibull is treated at the end of the next section on the basis of a numerical strategy that is described in Appendix B.

9.2 CLASSIFICATION OF DISTRIBUTIONS AND BOUNDS ON RENEWAL MEASURES

The next topic we should consider as a foundation for our examination of repairable systems is a reclassification of the basic life distributions and some associated bounds on the numbers of renewals. The idea here is that the length of each of the stochastically identical operating intervals is well represented by one of the life distributions with which we are already familiar. We now describe those distributions in terms that are meaningful for a series of intervals. To start, we compare life lengths of new and used devices.

Defn. 9.3: A life distribution, $F_T(t)$, is said to be *New Better than Used* (NBU) if

$$\bar{F}_T(t+u) \le \bar{F}_T(t)\bar{F}_T(u) \tag{9.19}$$

and to be *New Worse than Used* (NWU) if

$$\bar{F}_T(t+u) \ge \bar{F}_T(t)\bar{F}_T(u) \tag{9.20}$$

In other words, an NBU component has a life distribution that assigns a higher probability of survival for a fixed time to a new device than it does for a device that has already been used. An algebraic equivalent to Expression 9.19 is

$$\bar{F}_T(t+u \mid t) = \Pr[T \ge t+u \mid T > t] = \frac{\bar{F}_T(t+u)}{\bar{F}_T(t)} \le \bar{F}_T(u) \tag{9.21}$$

For an NWU device, the direction of the inequality is reversed.

As with our original classification of life distributions, a weaker classification than NBU and NWU has also been defined. This is,

Defn. 9.4: A life distribution, $F_T(t)$, is said to be *New Better than Used in Expectation* (NBUE) if

$$\int_t^\infty \bar{F}_T(u)du \le E_F[T]\bar{F}_T(t) \tag{9.22}$$

and to be *New Worse than Used in Expectation* (NWUE) if

$$\int_t^\infty \bar{F}_T(u)du \geq E_F[T]\bar{F}_T(t) \tag{9.23}$$

This definition may also be considered consistent with intuition, as rearranging the condition in Expression 9.22 yields

$$\frac{1}{\bar{F}_T(t)}\int_t^\infty \bar{F}_T(u)du \leq E_F[T] \tag{9.24}$$

for NBUE, and the inequality is reversed for an NWUE component. Expression 9.24 says that the mean of the residual life distribution for a used component is shorter than the mean of the life distribution for a new component.

Two final points concerning classifications of distributions are (1) that there are numerous additional classifications that are not treated here, and (2) the classifications that we have defined yield a natural nested ordering that can be useful. An example of a classification not treated here is increasing mean residual life (IMRL). The nested ordering is IFR $\subseteq$ IFRA $\subseteq$ NBU $\subseteq$ NBUE and DFR $\subseteq$ DFRA $\subseteq$ NWU $\subseteq$ NWUE.

The primary reason we define the classes of life distributions is that the classifications permit us to define bounds on renewal behaviors that we would not otherwise be able to obtain. Keep in mind the fact that, when we model the failure and renewal behavior of a device with a distribution other than the exponential, the renewal expressions may become intractable. In some such cases, we may use the exponential forms to obtain various bounds. The most commonly used examples of these bounds are:

a. If $F_T(t)$ is IFR and has finite mean $E_F[T]$, then for $0 \leq t < E_F[T]$,

$$\bar{F}_T(t) \geq e^{-t/E_F[T]} \tag{9.25}$$

b. If $F_T(t)$ is IFR and has finite mean $E_F[T]$, then for $0 \leq t < E_F[T]$,

$$F_T^{(n)}(t) \le 1 - \sum_{j=0}^{n-1} \frac{\left(t/E_F[T] \right)^j}{j!} e^{-t/E_F[T]} \tag{9.26}$$

c. If $F_T(t)$ is IFR and has finite mean $E_F[T]$, then for $0 \le t < E_F[T]$,

$$\Pr[N_t \ge n] \le \sum_{j=n}^{\infty} \frac{\left(t/E_F[T] \right)^j}{j!} e^{-t/E_F[T]} \tag{9.27}$$

d. If $F_T(t)$ is NBU and has cumulative hazard function $Z_T(t)$, then for $t \ge 0$,

$$\Pr[N_t < n] \ge \sum_{j=0}^{n-1} \frac{\left(Z_T(t) \right)^j}{j!} e^{-Z_T(t)} \tag{9.28}$$

e. If $F_T(t)$ is IFR and has cumulative hazard function $Z_T(t)$, then for $t \ge 0$,

$$\Pr[N_t < n] \le \sum_{j=0}^{n-1} \frac{\left(nZ_T(t/n) \right)^j}{j!} e^{-nZ_T(t/n)} \tag{9.29}$$

f. If $F_T(t)$ is NWU and has cumulative hazard function $Z_T(t)$, then for $t \ge 0$,

$$\Pr[N_t < n] \le \sum_{j=0}^{n-1} \frac{\left(Z_T(t) \right)^j}{j!} e^{-Z_T(t)} \tag{9.30}$$

g. If $F_T(t)$ is DFR and has cumulative hazard function $Z_T(t)$, then for $t \ge 0$,

$$\Pr[N_t < n] \ge \sum_{j=0}^{n-1} \frac{\left(nZ_T(t/n) \right)^j}{j!} e^{-nZ_T(t/n)} \tag{9.31}$$

h. If $F_T(t)$ is NBUE and has finite mean $E_F[T]$, then for $t \geq 0$,

$$M_{F_T}(t) \leq t \big/ E_F[T] \tag{9.32}$$

i. If $F_T(t)$ is NWUE and has finite mean $E_F[T]$, then for $t \geq 0$,

$$M_{F_T}(t) \geq t \big/ E_F[T] \tag{9.33}$$

Consider the interpretations and some examples of the above set of conditions.

Suppose we are studying the behavior of a population of components for which the Weibull distribution provides a good model of the dispersion in life lengths. Suppose further that we have estimated the parameters of that Weibull distribution to be $\beta = 2.75$ and $\theta = 4000$ hours. For this IFR distribution, the mean life is $E_F[T] = 3559.43$ hours, so the reciprocal of the mean life is 2.809×10^{-4}, and Expression 9.25 states that for $t \leq 3559.43$

$$\bar{F}_T(t) \geq e^{-t/3559.43} = e^{-0.0002809t}$$

so

$$\bar{F}_T(1000) \geq e^{-(0.0002809)(1000)} = 0.755$$

In this case, we can compute the reliability at 1000 hours to be 0.978. Perhaps more interesting is the fact that the probability that the third component life length is completed by 2000 hours is bounded by Expression 9.26 as

$$F_T^{(3)}(2000) \leq 1 - \sum_{j=0}^{2} \frac{\left(t/3559.43\right)^j}{j!} e^{-t/3559.43}$$

$$= 1 - (0.570 + 0.320 + 0.090) = 0.0195$$

On the other hand, Expression 9.27 indicates that the probability of three or more renewals in 2000 hours is bounded by the same quantity. That is,

$$\Pr[N_{2000} \geq 3] \leq 0.0195$$

For the same device population, Expressions 9.28 and 9.29 indicate that

$$0.9805 = \sum_{j=0}^{2} \frac{\left(Z_T(t)\right)^j}{j!} e^{-Z_T(t)} \leq \Pr[N_{2000} < 3]$$

$$\leq \sum_{j=0}^{2} \frac{\left(3Z_T(2000/3)\right)^j}{j!} e^{-3Z_T(2000/3)} = 0.999$$

Finally, for the same population, Expression 9.32 states that

$$M_{F_T}(t) \leq t/E_F[T] = t/3559.43$$

so

$$M_{F_T}(2000) \leq 2000/3559.43 = 0.562$$

Note further that the Expression 9.32 can be combined with the first of the Expressions of 9.18 to yield

$$t/E_F[T] - 1 \leq M_{F_T}(t) \leq t/E_F[T]$$

so

$$0.686 \leq M_{F_T}(6000) \leq 1.686$$

Naturally, we can perform similar example computations for Expressions 9.30, 9.31, and 9.33. If our Weibull population had a shape parameter of $\beta = 0.75$ rather than 2.75, the mean life would be 4762.56 hours, and Expression 9.30 indicates that

$$\Pr[N_{2000} < 3] \le 0.977$$

and

$$\Pr[N_{6000} < 3] \le 0.844$$

On the other hand, Expression 9.31 indicates that

$$\Pr[N_{2000} < 3] \le 0.998$$

and

$$\Pr[N_{6000} < 3] \le 0.735$$

In the case of Expression 9.33, we obtain

$$M_{F_T}(2000) \ge 0.420$$

and

$$M_{F_T}(6000) \ge 1.260$$

These examples illustrate the information one can obtain in cases in which the convolutions or renewal equations are computationally difficult.

As noted above, modern computing power has made previously taxing computations much more manageable. For example, the renewal function for the Weibull distribution cannot be expressed in closed form at all. However, Lomnicki [49] defined an equivalent infinite series expansion for the Weibull renewal function. The series is exact until it is truncated to finitely many terms, in which case it provides an approximation that is often quite accurate. The series form is provided in Appendix B. When the series is truncated at 15 terms and the value of the shape parameter is β = 2.75, we obtain $M_{F_T}(1000) = 0.022$, $M_{F_T}(2000) = 0.140$, $M_{F_T}(6000) = 1.272$, and $M_{F_T}(8000) = 1.821$. For the case in which β = 0.75, we obtain $M_{F_T}(1000) = 0.370$, $M_{F_T}(2000) = 0.641$, $M_{F_T}(6000) = 1.585$, and $M_{F_T}(8000) = 2.029$.

9.3 RESIDUAL LIFE DISTRIBUTION

In some cases, we wish to examine the behavior of a device after it has been operating for some period of time. The question of how much life remains to the device is a key question and is reasonable within the context of the life distribution as well as within the context of a renewal process. It is often posed relative to a renewal process. In addition, the residual life distribution can also be used to define bounds on renewals.

When examining a renewal process, we might reasonably ask the question of how long it is likely to be until the next renewal, in view of the fact that the previous renewal took place a particular number of hours ago. This is the same question as how long will a device of a given age continue to function.

For a device having a known age, say τ, the residual life distribution is the probability distribution on the longevity of the device from age τ onwards. We may consider the residual life distribution as the conditional distribution on additional life length given survival to age τ. Thus, if we denote the residual life distribution for a device of age τ by $F_{U(\tau)}(u)$, we can say that

$$\Pr[U(\tau) \le u] = F_{U(\tau)}(u) = \frac{F_T(u+\tau) - F_T(\tau)}{\bar{F}_T(\tau)} = \frac{\bar{F}_T(\tau) - \bar{F}_T(u+\tau)}{\bar{F}_T(\tau)} = 1 - \frac{\bar{F}_T(u+\tau)}{\bar{F}_T(\tau)} \tag{9.34}$$

and the corresponding reliability statement is

$$\Pr[U(\tau) > u] = \bar{F}_{U(\tau)}(u) = \frac{\bar{F}_T(u+\tau)}{\bar{F}_T(\tau)} \tag{9.35}$$

Of course, in both of these expressions, $F_T(t)$ is the underlying life distribution.

Consider two examples. First, if the underlying life distribution is exponential, the memoryless property of the distribution surfaces again, and we find that the residual life distribution is the same as the life distribution. That is,

$$F_{U(\tau)}(u) = 1 - \frac{\bar{F}_T(u+\tau)}{\bar{F}_T(\tau)} = 1 - \frac{e^{-\lambda(u+\tau)}}{e^{-\lambda\tau}} = 1 - e^{-\lambda u}$$

If the underlying life distribution is Weibull with shape parameter β = 2.75 and scale parameter θ = 4000 hours, then at 500 hours after a renewal, the distribution on the time to the next renewal is

$$F_{U(\tau=500)}(u) = 1 - \frac{\bar{F}_T(u+\tau)}{\bar{F}_T(\tau)} = 1 - \frac{e^{-\left(\frac{u+500}{2500}\right)^{2.75}}}{e^{-\left(\frac{500}{2500}\right)^{2.75}}} = 1 - 1.012e^{-\left(\frac{u}{2500}+0.2\right)^{2.75}}$$

so $F_{U(\tau=500)}(1000) = 0.208$ and $F_{U(\tau=500)}(2500) = 0.806$. The corresponding reliability values are the complements of the failure probabilities.

Here is another way to look at the residual life distribution. If we consider the progress of a renewal process over the time domain, the age, say τ, of a functioning device at any point in time, say t, is a random variable. The value of the random variable, device age, may be represented by

$$\tau = t - S_{N_t} \tag{9.36}$$

and the time until the next renewal is

$$u(\tau) = S_{N_t+1} - t \tag{9.37}$$

Given these definitions, we can represent the probability that, at any point in time, t, the residual life exceeds a specific value, u, as the probability that the first device in the process survives beyond $t + u$ plus the probability that a renewal occurred at some point in time prior to t and the device started at that time survives longer than $t + u$. That is,

$$\bar{F}_{U(\tau)}(u) = \Pr[U(\tau) > u] = \bar{F}_T(t+u) + \int_0^t \bar{F}_T(t+u-x)m_{F_T}(x)dx \tag{9.38}$$

Notice the similarity of Expression 9.38 to the Key Renewal Theorem.

The approach to the analysis of Expression 9.38 is the same as that for the Key Renewal Theorem and can be just as taxing. Nevertheless, there are some useful results we can obtain. First, the limiting form of the residual life distribution at any point in time is

$$F_U(u) = \frac{1}{E_F[T]} \int_0^u \bar{F}_T(x)dx \qquad (9.39)$$

Note that this form does not depend upon the age of the operating device. In general, as indicated in Chapter 8, Cox [44] has shown that the mean of this distribution will be

$$E_U[U] = \frac{E_F^2[T] + Var_F[T]}{2E_F[T]} \qquad (9.40)$$

In addition, we have the additional conditions that:

a. If $F_T(t)$ is NBU then $\bar{F}_U(u) \le \bar{F}_T(u)$
b. If $F_T(t)$ is NWU then $\bar{F}_U(u) \ge \bar{F}_T(u)$

Here again, these apply at any point in time.

To close this discussion, consider the following points. First, looking forward from any arbitrary point in time, the time until the next renewal is called the forward recurrence time. We have just obtained the primary results related to that quantity. Second, from any point in time, we may also look back and ask how long has it been since the most recent renewal. This quantity is called the backwards recurrence time. Cox [44] has shown that the forward recurrence time and the backward recurrence time have the same distributions, so the results described here apply to both. Finally, the key point here is that it is the use of the renewal density in Expression 9.38 that permits us to analyze the residual life (and recurrence times).

9.4 CONCLUSION

The models presented in this chapter serve to highlight the questions one should consider in the study of repairable systems. For some systems, it is the periods of operation that are the greatest concern, and the duration of repair is either negligible or unimportant. For those systems, treating repair as instantaneous is appropriate. Similarly, the study of individual components and the analysis of some systems may reasonably be based on renewal processes. Finally, the operating performance of many systems is improved by the use of preventive maintenance, while for some other systems, preventive maintenance is unproductive or impossible. As we study specific equipment items, we should tailor our models in terms of these operating features. The models presented in this chapter apply to the instantaneous repair case with renewal. More important, they establish the basic approaches to model formulation and analysis and emphasize the choices we must make.

9.5 EXERCISES

1. Consider a renewal process in which times between failures have a Normal distribution with a mean of 400 hours and a standard deviation of 50 hours. Construct the functional form for $F_{S_k}(t) = F_T^{(k)}(t)$ and the specific realization of this function when k = 3. Compute the mean and standard deviation for the distribution on S_3.
2. For the distribution on S_3 obtained in the preceding problem, compute the bounds defined by Expressions 9.18, 9.25, 9.26, 9.27, 9.28, 9.29, and 9.32.
3. Prove that an IFR distribution is NBU.
4. Consider a Weibull distribution for which β = 0.60 and θ = 1000 hours. For this distribution, compute the bounds defined by Expressions 9.18, 9.30, 9.31, and 9.33.
5. For the distribution of the preceding problem, use Lomnicki's method to obtain approximate values for the renewal function at 1000, 2000, and 5000 hours.
6. Let $F_T(t)$ be a Weibull distribution with β = 1.75 and θ = 800 hours. For this distribution, compute the

bounds defined by Expressions 9.18, 9.25, 9.26, 9.27, 9.28, 9.29, and 9.32.

7. For the distribution of the preceding problem, use Lomnicki's method to obtain approximate values for the renewal function at 2000, 4000, and 5000 hours.
8. Let $F_T(t)$ be an NBU distribution, and let $u(\tau)$ be the residual life at age τ of a component having life distribution $F_T(t)$. Show that $E_U[u(\tau)] \leq E_F[T]$, where $E_F[T]$ is the mean of $F_T(t)$.
9. Let $F_T(t)$ be a Weibull distribution with $\beta = 1.75$ and $\theta = 800$ hours. For this distribution, compute the reliability at $u = 100$, $u = 1000$, and $u = 2500$ hours for devices that have achieved ages of $\tau = 500$ hours, $\tau = 2000$ hours, and $\tau = 4000$ hours.

10

Repairable Systems II — Nonrenewal and Instantaneous Repair

As indicated in the previous chapter, there are many types of devices for which repair implies renewal. For those cases in which we are studying individual components in a specific equipment “slot,” treating component repair as a component renewal point is clearly appropriate. For some other systems, repair corresponds either exactly or approximately to system renewal, so the models described in the preceding chapter provide a reasonable portrayal of operating behavior.

On the other hand, there are many types of devices for which a repair does not return the unit to a new condition. There are also large complex systems, such as automobiles, for which the replacement of a few of its many components does not appreciably change the “age” of the system. For equipment of this sort, the unit is not as good as new following repair, so unit age following repair may not be taken to be zero. Clearly, the state of a device following its repair determines whether or not a sequence of operating periods is well modeled by a common distribution. When system age following repair is nonzero, successive operating periods do not have a common distribution, and the renewal model does not apply.

Several models based on nonstationary processes have been suggested for those devices that are not renewed by repair. We shall explore some of them here. While doing so, we will continue to assume that repair is instantaneous, or equivalently, that the operating periods are our key concern and repair intervals are negligible or unimportant.

The idea of a nonstationary process is that we have a sequence or operating intervals, each of which ends with unit failure. As in the case of the renewal process, we denote the lengths of the intervals by T_1, T_2, T_3, …, but for the nonstationary process, each interval has a distinct distribution. Three useful models have been developed to treat the nonstationary sequence of operating intervals.

10.1 MINIMAL REPAIR MODELS

The earliest model of system behavior for the case in which repair does not imply renewal was suggested by Barlow and Hunter [50]. Their model is called the "minimal repair" model because it is constructed by assuming that when a failure of a device occurs, the unit is repaired and is placed back in operation without any change in its age. That is, the repaired unit has the same age and the same failure hazard following repair as it had just prior to failure. It is "as bad as old." The algebraic form of this model follows directly the construction of a nonstationary process. In fact, it is the nonhomogeneous Poisson process (NHPP) model. The construction starts with the time process. We denote the number of failures over the interval $(0, t)$ by N_t and let

$$\Lambda(t) = E[N_t] \tag{10.1}$$

represent its expected value. We also define the failure intensity function, $\lambda(t)$ so that

$$\Lambda(t) = \int_0^t \lambda(u)du \tag{10.2}$$

The definition of the accumulating operating time, S_k, is the same as defined in Expression 9.1, and the basic relationship between the variables in the time domain and those in the frequency domain, Expression 9.4, still applies. Therefore, we can say that in the frequency domain

$$\Pr[N_{t+x} - N_t = k] = \frac{\left(\Lambda(t+x) - \Lambda(t)\right)^k}{k!} e^{-\left(\Lambda(t+x)-\Lambda(t)\right)} \qquad (10.3)$$

and in the time domain

$$\Pr[T_{k+1} - T_k > t] = e^{-\left(\Lambda(T_k+t)-\Lambda(T_k)\right)} \qquad (10.4)$$

In many respects, this is a very appealing model. It is relatively easy to use and to analyze. One need only specify a failure intensity function, and this is reasonably obvious, as the choice must be consistent for T_1. So by implication,

$$\Lambda(t) = Z_T(t) \qquad (10.5)$$

which is to say, the failure intensity function must correspond to the hazard function of the device life distribution.

Reversing this logic, if the failure intensity corresponds to the device hazard function, then the NHPP model provides the appropriate representation of system behavior. In this case, Equations 10.3 and 10.4 yield the system performance information of interest.

For an example, suppose we have a device for which $\Lambda(t) = (t/\theta)^\beta$, with θ = 4000 hours and β = 2.75. Then $\Lambda(2000) = 0.149$, $\Lambda(4000) = 1.0$, $\Lambda(6000) = 3.049$, and $\Lambda(8000) = 6.727$. Also, using Expression 10.3, we obtain results such as

$$\Pr[N_{2000} = 2] = e^{-\left(\Lambda(2000)-\Lambda(0)\right)} \frac{\left(\Lambda(2000) - \Lambda(0)\right)^2}{2!}$$

$$= 0.0095$$

$$\Pr[N_{2000} = 1] = e^{-\left(\Lambda(2000)-\Lambda(0)\right)} \frac{\left(\Lambda(2000) - \Lambda(0)\right)}{1!}$$

$$= 0.1281$$

$$\Pr[N_{2000} = 0] = e^{-\left(\Lambda(2000)-\Lambda(0)\right)} \frac{\left(\Lambda(2000) - \Lambda(0)\right)^0}{0!}$$

$$= 0.8619$$

$$\Pr[N_{4000} - N_{2000} = 2] = e^{-\left(\Lambda(4000)-\Lambda(2000)\right)} \frac{\left(\Lambda(4000) - \Lambda(2000)\right)^2}{2!}$$

$$= 0.1547$$

$$\Pr[N_{4000} - N_{2000} = 1] = e^{-\left(\Lambda(4000)-\Lambda(2000)\right)} \frac{\left(\Lambda(4000) - \Lambda(2000)\right)^1}{1!}$$

$$= 0.3634$$

$$\Pr[N_{4000} - N_{2000} = 0] = e^{-\left(\Lambda(4000)-\Lambda(2000)\right)} \frac{\left(\Lambda(4000) - \Lambda(2000)\right)^0}{0!}$$

$$= 0.4268$$

$$\Pr[N_{6000} - N_{4000} = 2] = e^{-\left(\Lambda(6000)-\Lambda(4000)\right)} \frac{\left(\Lambda(6000) - \Lambda(4000)\right)^2}{2!}$$

$$= 0.2705$$

$$\Pr[N_{6000} - N_{4000} = 1] = e^{-\left(\Lambda(6000)-\Lambda(4000)\right)} \frac{\left(\Lambda(6000) - \Lambda(4000)\right)^1}{1!}$$

$$= 0.2640$$

$$\Pr[N_{6000} - N_{4000} = 0] = e^{-\left(\Lambda(6000)-\Lambda(4000)\right)} \frac{\left(\Lambda(6000) - \Lambda(4000)\right)^0}{0!}$$

$$= 0.1288$$

and

$$\Pr[N_{8000} - N_{6000} \leq 2] = e^{-\left(\Lambda(8000)-\Lambda(6000)\right)} \sum_{k=0}^{2} \frac{\left(\Lambda(8000) - \Lambda(6000)\right)^k}{k!}$$

$$= 0.2892$$

Similarly, using Expression 10.4, we can compute

$$\Pr[T_2 - T_1 > 2000] = e^{-(\Lambda(T_1+2000)-\Lambda(T_1))}$$

so

$$T_1 = 500 \rightarrow \Pr[T_2 - T_1 > 2000] = e^{-(\Lambda(2500)-\Lambda(500))} = 0.7624$$

$$T_1 = 1500 \rightarrow \Pr[T_2 - T_1 > 2000] = e^{-(\Lambda(3500)-\Lambda(1500))} = 0.5351$$

$$T_1 = 3500 \rightarrow \Pr[T_2 - T_1 > 2000] = e^{-(\Lambda(5500)-\Lambda(3500))} = 0.1812$$

and

$$\Pr[T_1 > 2000] = e^{-(\Lambda(2000))} = 0.8619$$

Note the implied decline in the 2000-hour reliability with increasing reference time. This is a feature identified as indicative of increasing hazard in Chapter 4.

The NHPP also has the feature that it can be transformed into an equivalent homogeneous Poisson process. The transformation is basically a revision in the time scale. Suppose we are considering a nonhomogeneous process with cumulative intensity function $\Lambda(t) = E[N_t]$, as stated in Expression 10.1. Let

$$\tau(t) = \inf\{s \mid \Lambda(s) > t\} \tag{10.6}$$

which is to say that τ is the time inverse for the cumulative intensity function of the process. Then, we define the process:

$$N^s_{\tau(t)} = N_t \tag{10.7}$$

and we find that $N^s_{\tau(t)}$ is a homogeneous Poisson process with intensity equal to 1. That is, if time is measured using the scale $\tau(t)$, then an observed sequence of failures will appear to constitute a stationary Poisson process with intensity equal to 1. In fact, it is the correspondence between the homogeneous and the nonhomogeneous processes that permits us to

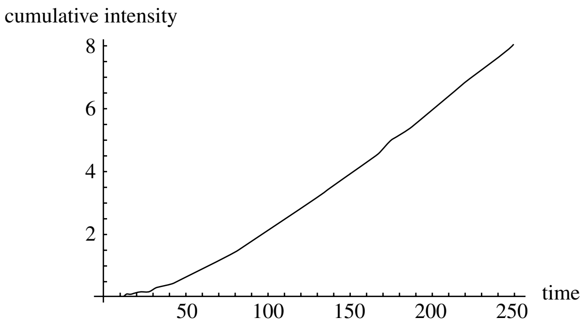


Figure 10.1 Cumulative intensity function for an NHPP based on a Gamma distribution.

define the probabilities on the number of events in Expression 10.3. The real utility of the relation is that we can sometimes use our results for renewal processes to obtain probabilities and other information for nonhomogeneous processes such as the one used to represent minimal repair.

Consider an example. Suppose we are interested in a device for which we believe the life distribution is Gamma, with parameters $\lambda = 0.05$ and $\beta = 3$, and for which minimal repair is assumed. In this case, using Expression 10.5, we obtain a plot of the cumulative intensity function as shown in Figure 10.1. Reading from the plot, we note that $\Lambda(100) = 2.082$. Thus, we can compute the probabilities for the possible number of failures over the first 100 hours of device operation. These are shown in Table 10.1. We can also determine that the time between the second and third failures has an expectation of about 28.5 hours.

10.2 IMPERFECT REPAIR MODELS

The minimal repair models were the first to capture the feature of real operating equipment that often is not renewed by repair. They were thus the first of the nonstationary models and provided an advance in realism. On the other hand, most devices do not remain "as bad as old" when repaired. Actual device behavior is often somewhere between the extremes of

Table 10.1 Frequency Probabilities for an NHPP Based on the Equivalent HPP

N_{100}	Pr
0	0.125
1	0.260
2	0.270
3	0.186
4	0.098
5	0.041
6	0.014
7	0.004
8	0.001

renewal and simple continuation. In an attempt to make the nonstationary models more realistic, Brown and Proschan [51] defined "imperfect repair." Under an imperfect repair regime, following repair, a device is renewed (as good as new) with probability p and minimally repaired (as bad as old) with probability $q = 1 - p$. The result is that the probability process that represents equipment experiences is a mixture of the renewal process and the minimal repair NHPP.

The key to the analysis of this basic imperfect repair model is the fact that the times at which a perfect repair is performed constitute a renewal process that is embedded in the more general point process produced by the device failures. The points at which a perfect repair occurs are restart points. Brown and Proschan [51] point this out and then proceed with the analysis of the process as follows. Denote the distribution on the time interval between perfect repairs by $F_p(t)$ and its hazard function by $z_p(t)$. Now, if a device has life distribution $F_T(t)$ and only minimal repairs occur during $(0, t)$, then at time t, the device behaves as if it has age t. Its hazard function is $z_T(t)$, and if a failure occurs, the associated repair will be perfect with probability p. Thus, the conditional intensity of the occurrence of a perfect repair is

$$z_p(t) = pz_T(t) \tag{10.8}$$

As with any hazard function, the survivor function corresponding to $z_p(t)$ is

$$\bar{F}_p(t) = e^{-\int_0^t z_p(u)du} = e^{-p\int_0^t z_T(u)du} = e^{-pZ_T(t)}$$
$$= \left(e^{-Z_T(t)}\right)^p = \left(\bar{F}_T(t)\right)^p \tag{10.9}$$

Consequently, the distribution on the duration of the intervals between perfect repairs, which is to say renewals, is

$$F_p(t) = 1 - \left(\bar{F}_T(t)\right)^p = 1 - e^{-pZ_T(t)} \tag{10.10}$$

This is a very useful and very general result. It is useful first because it applies to all choices of underlying life distribution. In addition, Expression 10.8 and those that follow it show that the distribution on the time between perfect repairs will be of the same class as the life distribution — IFR, IFRA, DFR, and so on. Within the intervals between perfect repairs, Expressions 10.3 and 10.4 describe the probabilities on the frequency of minimal repairs and the times between them. Finally, all of the results for renewal processes that we have examined apply to the intervals between perfect repairs.

Consider a simple example. Suppose that failures of a particular device are well modeled by a Weibull distribution with parameters $\theta = 4000$ hours and $\beta = 2.75$. Suppose further that the probability of perfect repair is $p = 0.25$. Using these values, we compute the values for the life distribution as usual, and the calculation of the times between renewals has the distribution

$$F_p(t) = 1 - \left(\bar{F}_T(t)\right)^p = 1 - e^{-p\left(t/\theta\right)^\beta} = 1 - e^{-0.25\left(t/4000\right)^{2.75}}$$

The results of these calculations are plotted in Figure 10.2. Note that the distribution on individual life lengths is stochastically smaller than the distribution on renewal times. Naturally, as p is increased, the times between perfect repairs tend to be shorted, and as p is decreased, the times tend to become longer.

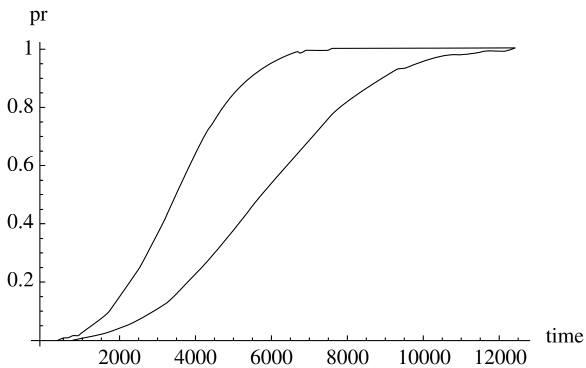


Figure 10.2 Example life and renewal time distributions under imperfect repair.

Given the apparent utility of the above imperfect repair model, Block, Borges, and Savits [52] suggested that the models could be enhanced by making p age dependent. They define the probability of a perfect repair for a device of age t to be $p(t)$, with the complementary probability of $q(t) = 1 - p(t)$ for a minimal repair. In a sequence of failures with minimal repairs, the device is considered to continue to age, so the time since the last perfect repair is the age of the device. Using logic that is similar to but rather more intricate than that for the simple (p, q) model, Block, Borges, and Savits [52] show that the distribution on the times between perfect repairs has the same basic form as that for the case of age-independent perfect repair probabilities. That is,

$$z_p(t) = p(t)z_T(t) = \frac{p(t)f_T(t)}{\bar{F}_T(t)} \tag{10.11}$$

leads to the distribution

$$F_p(t) = 1 - e^{-Z_p(t)} = 1 - e^{-\int_0^t \frac{p(u)f_T(u)}{\bar{F}_T(u)}du} \tag{10.12}$$

As in the simpler case, the times of perfect repairs form a renewal process, and it is again the case that we can compute

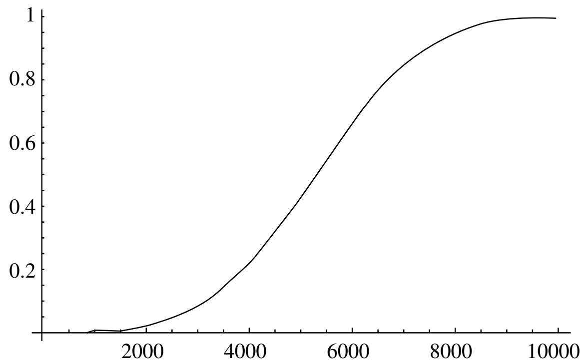


Figure 10.3 Example renewal time distributions under age-dependent imperfect repair.

device behavior measures using the methods discussed above. It is also often, but not always, the case that the distribution on times between renewals has the same behavior as the life distribution. Behavior is preserved for IFR, IFRA, NBU, DFR, DFRA, and NWU hazards but not for NBUE.

Consider an example device for which failures are well modeled by a Weibull distribution with parameters θ = 4000 hours and β = 2.75. Suppose further that the probability function for perfect repair is $p(t) = 1 - e^{-\rho t}$ where ρ = 0.0001. Using these parameter values, we find that the distribution on the time between perfect repairs (renewals) is the one shown in Figure 10.3. In addition, as noted above, during the intervals between renewals, we use the cumulative intensity function (Expression 10.5) of

$$\Lambda(t) = Z_p(t) \tag{10.13}$$

and Expressions 10.3 and 10.4 to compute specific event probabilities.

Clearly, the imperfect repair models permit us to move away from the "pure renewal" models of device behavior. However, the use of the mixture of repair types is rather artificial and does not really provide a true picture of operating experiences. More recent models using what are called equivalent age measures have been developed to try to provide more realistic representations of repair processes.

10.3 EQUIVALENT AGE MODELS

In order to provide a more realistic portrayal of equipment state following repair, two sets of authors have developed models that define an equivalent age following repair. Kijima [53, 54] defines two models for what he calls general repair. Wang and Pham [55, 56] define a "quasi-renewal" process model to represent postrepair device state. We shall examine all of these models here.

Before examining these models, recall that the residual life distribution is the distribution on the remaining life length for a device that has already operated for some time. In Chapter 9, we found that the residual life distribution for a device of age τ can be defined in terms of the device life distribution (Expression 9.34) as

$$\Pr[U(\tau) \le u] = F_{U(\tau)}(u) = 1 - \frac{\bar{F}_T(u+\tau)}{\bar{F}_T(\tau)}$$

Clearly, as with all distributions, the residual life distribution has a mean value, $E[U(\tau)]$, and if this mean value is decreasing in τ, we say the distribution is in the class of DMRL (decreasing mean residual life) distributions. All IFR distributions are DMRL, and the models developed by Kijima and by Wang and Pham apply specifically to cases in which the underlying life distribution of the device of interest is DMRL. The condition DMRL is weaker than IFR but implies device deterioration over time for all repair scenarios.

10.3.1 The Kijima Models

The Kijima I model is constructed as follows. Let T_j , j = 1, 2, … to represent the lengths of successive equipment operating intervals. Then assume that a repair, say the n^{th}, can ameliorate the damage or aging experienced by the equipment during only the most recent operating interval and not any of the damage or aging that was incurred during earlier intervals.

Thus, following the n^{th} repair, the virtual (or equivalent) age of the device, A_n, is defined to be

$$A_n = A_{n-1} + \pi_n T_n \tag{10.14}$$

In this expression, π_n is the fraction of the age that is accumulated during the n^{th} operating interval that is not "healed" by repair and is thus the additional age that is accumulated during the n^{th} interval. We may consider that $1 - \pi_n$ is the degree of repair for the n^{th} repair action. In general, we refer to the quantities π_n as "repair effectiveness factors." (Repair ineffectiveness factors might be more appropriate.)

The Kijima II model is similar. The difference is that the amelioration of accumulated damage or age is assumed to apply to all of the accumulated age. Thus, the virtual age of a device following the n^{th} repair is taken to be

$$A_n = \pi_n \left(A_{n-1} + T_n \right) \tag{10.15}$$

The selection of the values π_n determines the form of these models. In both cases, $0 \leq \pi_n \leq 1$ for all n. If we take $\pi_n = 0$ for all n, both Model I and Model II reduce to the renewal model. On the other hand, if we set $\pi_n = 1$ for all n in Model I, we obtain the minimal repair model. If we take π_n to be a Bernoulli random variable for each n in Model II, the result is the imperfect repair model. Because the various choices of coefficients lead to the previous models, Kijima called his models general repair models. He also specified that, in the most general case, the coefficients π_n should be taken to be random variables with any arbitrary and not necessarily identical distributions.

For all choices of the coefficients, the analysis of the Kijima models is complicated. In general, for any set of coefficients, the distribution on the duration of any operating interval can be defined for both Model I and Model II using the residual life distribution. The general form is

$$\Pr[T_n \leq t \,|\, A_{n-1} = u] = F_{T_n(u)}(t) = \frac{F_T(t+u) - F_T(u)}{\bar{F}_T(u)} \tag{10.16}$$

where, naturally, we take $A_0 = 0$. Note that the complement is

$$\Pr[T_n > t \mid A_{n-1} = u] = \frac{\bar{F}_T(t+u)}{\bar{F}_T(u)} \tag{10.17}$$

Now, we would like to be able to identify the probabilities and the expectations for the numbers of repairs over time and the time until a given number of repairs have been made. If we let $\Pi = \{\pi_1, \pi_2, ...\}$ represent the sequence of repair effectiveness factors, then we can identify algebraic representations for these measures. Specifically, we note that the sequence of random variables $\{A_n\}$ represents the "virtual age" stochastic process, and that the random variable

$$S_n = \sum_{j=1}^{n} T_j \tag{10.18}$$

is the real age (or elapsed time) for the sequence of operating intervals. We can denote the distribution on the real age by

$$F_{S_n}(t) = \Pr[S_n \le t] \tag{10.19}$$

and the number of repairs by

$$N_t = \sup\{n \ge 1 \mid S_n \le t\} \tag{10.20}$$

Then, the usual relationship between the time and frequency domains implies that

$$\Pr[N_t \ge n] = F_{S_n}(t) \tag{10.21}$$

and (in principle) we can compute expectations on the measures as

$$E[N_t] = \sum_{n=1}^{\infty} F_{S_n}(t) \tag{10.22}$$

and

$$E[S_n] = \int_0^{\infty} \bar{F}_{S_n}(u)du \tag{10.23}$$

Unfortunately, for the most general definitions of the vector Π, these measures of interest are extremely difficult to determine. As shown below, even for quite simple choices of Π, the exact analysis of Expressions 10.22 and 10.23 require successive numerical integrations that are intricate. Rather than attempt the difficult numerical analysis, we usually compute an upper bound on $E[S_n]$ and use that as our key measure of system behavior.

To appreciate the utility of the bounds, we examine the analysis of the basic models. Assume that the factors π_n are independently and identically distributed (i.i.d.) random variables having expected value $E[\pi]$. Start with Model I.

Suppose that $\pi_{n+1} = \pi$ for $\pi \in [0, 1]$ and assume that $\pi \neq 0$, so that the application of Expression 10.17 yields

$$\bar{F}_{A_{n+1}}(t) = \Pr[A_{n+1} > t] = \Pr[A_n > t]$$

$$+ \int_0^t \Pr[\pi T_n > t - u \mid A_n = u] \Pr[A_n = u] du$$

$$= \bar{F}_{A_n}(t) + \int_0^t \frac{\bar{F}_T\left(u + \frac{t-u}{\pi}\right)}{\bar{F}_T(u)} dF_{A_n}(u)$$

$$= \bar{F}_{A_n}(t) + \int_0^t \frac{\bar{F}_T\left(u + \frac{t-u}{\pi}\right)}{\bar{F}_T(u)} f_{A_n}(u) du \tag{10.24}$$

and the density function is

$$f_{A_{n+1}}(t) = \frac{1}{\pi} \int_0^t \frac{f_T\left(u + \frac{t-u}{\pi}\right)}{\bar{F}_T(u)} f_{A_n}(u) du \tag{10.25}$$

with

$$f_{A_1}(t) = \frac{1}{\pi} f_T\left(t/\pi\right) \tag{10.26}$$

Now that the density on A_n is defined, we can construct the expected value of each operating interval using Expression 10.17 and 10.25. Specifically,

$$\begin{aligned} E[T_n \mid A_{n-1} = u] &= \int_0^{\infty} \Pr[T_n > t \mid A_{n-1} = u]dt \\ &= \int_0^{\infty} \frac{\bar{F}_T(t+u)}{\bar{F}_T(u)} dt \end{aligned} \tag{10.27}$$

and

$$E[T_n] = \int_0^{\infty} E[T_n \mid A_{n-1} = u] f_{A_{n-1}}(u)du \tag{10.28}$$

Clearly, one possible approach to computing $E[S_n]$ is as the sum of the terms $E[T_j]$.

Now, Kijima and associates [54] suggest that Expression 10.24 be restated as

$$\bar{F}_{A_{n+1}}(t) = \int_0^{t} g_t(u) f_{A_n}(u)du \tag{10.29}$$

where

$$g_t(u) = \begin{cases} \bar{F}_{U(u)}(\dfrac{t-u}{\pi}) & u \le t \\ 1 & u > t \end{cases} \tag{10.30}$$

He then notes that when the elements of the vector of repair effectiveness factors are i.i.d. random variables having distribution function $F_{\Pi}(\pi)$, the stochastic process, $\{A_n\}$ having A_0 = 0 is a Markov process with transition probability function

$$\Pr[A_{n+1} \le t \mid A_n = u] = \int_0^{1} \left(1 - g_t(u)\right) f_{\Pi}(\pi)d\pi \tag{10.31}$$

Using this result and the assumption that the underlying life distribution is DMRL, Kijima shows that bounds for $E[S_n]$ can be obtained. A lower bound is obtained when it is assumed that $\pi_n = \pi$ for all n and π is simply a constant. In that case, successive evaluation of the integrals stated in Expressions 10.27 and 10.28 are demanding but possible. For an upper bound for $E[S_n]$, we use the corresponding quantity for the imperfect repair case. That is, taking the π_n to be Bernoulli with $E[\pi_n] = p$, we have the imperfect repair case. For that model, we can compute $E[S_n, p]$ using the following recursion:

$$v(m, n) = \mu_m + (1-p)v(m, n-1) + pv(m+1, n-1) \quad (10.32)$$

where $v(m,n)$ is the expected value of S_n for a device that is subject to an imperfect repair regime and has already had $m - 1$ minimal repairs. The quantity μ_m is the mean length of m intervals when only minimal repair is used, and p is the probability that repair is perfect ($\pi = 0$). The boundary conditions for the recursion are $v(m, 1) = \mu_m$, and the values $v(1, n) = E[S_n, p]$ provide the upper bounds.

The interpretation of Expression 10.32 is that the expected value of the sum of the lengths of the next n operating intervals for a device that has had m minimal repairs so far is the expected value length of the next operating interval, μ_m, plus the expected value of the sum of the lengths of the following $n - 1$ intervals, for which the number of minimal repairs will be $m + 1$, with probability p, and m, with probability $1 - p$. The upper bound value that we obtain using Expression 10.32 applies to the "general repair case" in which the distribution on the π_n has expected value $E[\pi_n] = p$.

Based on our understanding of the minimal repair case, we can compute the value of the mean length of m operating intervals under minimal repair as

$$\mu_m = \frac{1}{\Gamma(m)} \int_0^\infty \left(Z_T(u)\right)^{m-1} e^{-Z_T(u)} du \quad (10.33)$$

This quantity forms the basis for the numerical computations for the recursion of Expression 10.32.

Consider an example. Suppose the underlying life distribution for a device is Weibull with $\beta = 2.75$ and $\theta = 4000$ hours. In this case, we calculate

Table 10.2 Values of Mean Residual Life under Minimal Repair

n	μ_n
1	3559.43
2	1294.34
3	882.50
4	695.31
5	584.69
6	510.27
7	456.15
8	414.69
9	381.70
10	354.71

$$Z_T(t) = \left(t/\theta \right)^{\beta} = \left(t/4000 \right)^{2.75}$$

so we easily obtain the values shown in Table 10.2.

Notice the clearly decreasing values of these mean residual life lengths. Using the values of Table 10.2, we can solve the recursion equation of Expression 10.32 to obtain the values of $E[S_n, p]$ which are our upper bounds on $E[S_n]$ for the general repair model. Example results are shown in Table 10.3. As indicated previously, the nested integrals make the calculation of the lower bounds very difficult. Using Expressions 10.28 and 10.25, the calculation of $E[T_2]$ is manageable. We find that for π = 0.50, $E[T_2] = 760.168$, for π = 0.75, $E[T_2] = 591.923$, and for π = 0.90, $E[T_2] = 514.559$. The corresponding lower bounds on $E[S_2]$ are obtained by adding the known value of $E[T_1] = 3559.43$ to these quantities to obtain $E[S_2] \geq 4319.60$, $E[S_2] \geq 4151.35$, and $E[S_2] \geq 4073.99$, respectively.

The extension of our analysis to the Kijima II model is reasonably direct. We start with the basic model stated in Expression 10.15 and assume the π_n are i.i.d. random variables. We then use Expression 10.17 again to obtain

Table 10.3 Values of $E[S_n, p]$ for Minimal Repair as Upper Bounds on $E[S_n]$ for General Repair

n	$E[\pi_n] = p = 0.90$	$E[\pi_n] = p = 0.75$	$E[\pi_n] = p = 0.50$
1	3559.43	3559.43	3559.43
2	5080.27	5420.04	5936.81
3	6063.38	6624.28	7743.96
4	6823.51	7527.55	9092.12
5	7457.78	8266.61	10179.47
6	8009.35	8902.29	11093.31
7	8501.38	9465.85	11886.39
8	8948.05	9975.54	12591.67
9	9358.77	10443.04	13230.49
10	9740.17	10876.37	13817.18

$$\begin{aligned}\bar{F}_{A_{n+1}}(t) &= \Pr[A_{n+1} > t] = \Pr[A_n > t/\pi] \\ &\quad + \int_0^{t/\pi} \Pr[T_n > t/\pi - u \mid A_n = u]\Pr[A_n = u]du \\ &= \bar{F}_{A_n}(t/\pi) + \int_0^{t/\pi} \frac{\bar{F}_T\left(t/\pi\right)}{\bar{F}_T(u)} dF_{A_n}(u) = \bar{F}_{A_n}(t) \\ &\quad + \int_0^{t} \frac{\bar{F}_T\left(t/\pi\right)}{\bar{F}_T(u)} f_{A_n}(u)du\end{aligned} \tag{10.34}$$

and the density function is

$$f_{A_{n+1}}(t) = \frac{1}{\pi} f_T\left(t/\pi\right) \int_0^{t/\pi} \frac{f_{A_n}(u)}{\bar{F}_T(u)} du \tag{10.35}$$

with

$$f_{A_1}(t) = \frac{1}{\pi} f_T(t/\pi) \tag{10.36}$$

as before. The general statements of the expected values for the S_n of Expressions 10.27 and 10.28 are the same for Model II as for Model I. Their analysis is no less complicated for Model II, so our approach is the same.

To compute upper bounds for the $E[S_n]$, we again use the corresponding quantity for the imperfect repair case. We take the π_n to be Bernoulli with $E[\pi_n] = p$, and we compute $E[S_n, p]$ using the following recursion:

$$v(m, n) = \mu_m + (1 - p)v(1, n - 1) + pv(m + 1, n - 1) \quad (10.37)$$

where the values $v(1, n) = E[S_n, p]$ are the upper bounds on $E[S_n]$ for the general repair case when the π_n are i.i.d., and the distribution on the π_n has expected value $E[\pi_n] = p$.

For Model II, we also note that repeated substitution within the recursion Expression 10.37 leads to the relation that, under imperfect repair, when $n \geq 2$,

$$E[T_n] = E[S_n] - E[S_{n-1}] = (1 - p)\sum_{j=1}^{n-1} p^{j-1}\mu_j + p^{n-1}\mu_n \quad (10.38)$$

The quantities μ_n in each of the above expressions are the same as for Model I. They are the successive mean residual life lengths for a device subjected to a minimal repair regime.

Consider the same example as above. When the life distribution is Weibull with $\beta = 2.75$ and $\theta = 4000.0$, the mean residual life values are those listed in Table 10.2, and the upper bounds on $E[S_n]$ defined in Expression 10.37 are shown in Table 10.4. Here again, the computational effort associated with the computation of the lower bounds is excessive.

In summary, we may observe that the Kijima models are very appealing, because they provide a significantly more realistic image of the state of equipment following repair. Unfortunately, the models are correspondingly difficult to analyze. On the other hand, it is not difficult to simulate the models and to use the simulation output to describe system behavior. This approach is now widely used.

Table 10.4 Values of $E[S_n, p]$ for Minimal Repair as Upper Bounds on $E[S_n]$ for General Repair

n	$E[\pi_n] = p = 0.90$	$E[\pi_n] = p = 0.75$	$E[\pi_n] = p = 0.50$
1	3559.43	3559.43	3559.43
2	5080.27	5420.04	5986.31
3	6267.53	7048.99	8310.23
4	7318.33	8598.97	10610.76
5	8296.54	10113.95	12904.37
6	9230.82	11611.27	15195.65
7	10136.33	13098.95	17486.09
8	11022.02	14581.11	19776.21
9	11893.50	16059.96	22066.19
10	12754.52	17536.78	24356.13

10.3.2 The Quasi-Renewal Process

An alternative approach to modeling the postrepair state of a device is suggested by Wang and Pham [56]. Rather than directly adjust the age of the unit following repair, Wang and Pham adjust the life distribution. For the sequence of operating intervals, $T_1, T_2, T_3, \ldots$, each of which ends with device failure and repair, they assume that

$$T_n = \alpha^{n-1} X_n \tag{10.39}$$

where α (>0) is a constant that alters the scale of the distribution and the X_n are i.i.d. random variables. Under this definition, the sequence $\{T_n\}$ is said to form a quasi-renewal process. Clearly, if $\alpha = 1$, the sequence is a renewal process. An interesting feature of this model is that a choice of $\alpha < 1$ implies that the operating intervals are decreasing in magnitude, as might occur with aging and deterioration. On the other hand, a choice of $\alpha > 1$ might be used to represent an ongoing improvement in the quality of replacement units, with a corresponding gradual increase in the duration of operating intervals.

For the models developed here, we assume that $\alpha < 1$, as the unit of interest is aging. The expressions developed apply to other choices of the parameter, but this is not pursued here. For any particular choice of α, we find that

$$\begin{aligned} F_{T_n}(t) &= F_X\left(t \big/ \alpha^{n-1}\right) \\ f_{T_n}(t) &= \frac{1}{\alpha^{n-1}} f_X\left(t \big/ \alpha^{n-1}\right) \\ E[T_n] &= \alpha^{n-1} E[X] \end{aligned} \tag{10.40}$$

Now, the similarity to the renewal process permits us to construct some descriptive expressions for the model. In particular, we again define S_n to be the sum of the T_j and N_t to be the number of failures during (0, t]. Then, in contrast to the Kijima models, we can immediately compute:

$$E[S_n] = \sum_{j=1}^{n} E[T_j] = E[X] \sum_{j=1}^{n} \alpha^{j-1} = \frac{1-\alpha^n}{1-\alpha} E[X] \tag{10.41}$$

In addition, we use the usual time frequency relationship:

$$\Pr[N_t \geq n] = \Pr[S_n \leq t]$$

and the same logic that we used for the renewal process to construct the quasi-renewal function:

$$M_{F_T}(t) = E[N_t] = \sum_{n=1}^{\infty} F_{S_n}(t) \tag{10.42}$$

Of course, the distributions $F_{S_n}(t)$ are all distinct, because they are convolutions of distinct distributions from a common class. However, since the process $\{T_n\}$ is constructed so regularly, the Laplace transforms for the distinct distributions and hence for the quasi-renewal function are readily constructed as

$$F^*_{T_n}(s) = F^*_X(\alpha^{n-1} s) \tag{10.43}$$

and

$$M^*_{F_T}(s) = \sum_{n=1}^{\infty} F^*_{S_n}(s) = \sum_{n=1}^{\infty} \prod_{j=1}^{n} F^*_X(\alpha^{n-1} s) \tag{10.44}$$

Also, as with the renewal process, the derivative of the quasi-renewal function is the quasi-renewal intensity function:

$$m_{F_T}^*(s) = \sum_{n=1}^{\infty} f_{S_n}^*(s) = \sum_{n=1}^{\infty} \prod_{j=1}^{n} f_X^*(\alpha^{n-1}s) \tag{10.45}$$

As in the case of the renewal model, it is usually quite difficult to obtain numerical values for the device performance measures. As indicated above, Expression 10.41 permits us to compute $E[S_n]$. On the other hand, the quasi-renewal function tends to be quite difficult.

Consider an example. If the life distribution is Normal with mean μ and standard deviation σ, then using the second of the relationships in (10.40), we find that each of the T_n has a Normal distribution:

$$f_{T_n}(t) = \frac{1}{\alpha^{n-1}} f_X\left(t/\alpha^{n-1}\right) = \frac{1}{\alpha^{n-1}\sqrt{2\pi\sigma^2}} e^{-\left(t/\alpha^{n-1} - \mu\right)^2 / 2\sigma^2}$$

$$= \frac{1}{\sqrt{2\pi(\alpha^{n-1}\sigma)^2}} e^{-\left(t - \mu\alpha^{n-1}\right)^2 / 2(\alpha^{n-1}\sigma)^2}$$

but with a mean value of $\alpha^{n-1}\mu$ and a standard deviation of $\alpha^{n-1}\sigma$. We can thus construct the convolutions of the distributions on the T_n to obtain the distributions on the S_n. We do this most easily using the Laplace transforms. For the normal life distribution,

$$f_X^*(s) = e^{-s\mu + \frac{s^2\sigma^2}{2}}$$

so

$$f_{T_n}^*(s) = e^{-s\alpha^{n-1}\mu + \frac{s^2(\alpha^{n-1}\sigma)^2}{2}}$$

and

Table 10.5 Values of $M(t)$ for the Quasi-Renewal Model with Normal Life Distribution

t	α = 0.95	α = 0.85	α = 0.75
1000	0.001	0.001	0.001
2500	0.067	0.067	0.067
5000	0.863	0.875	0.897
8000	1.577	1.741	1.992
10000	2.145	2.447	3.084

$$f^*_{S_n}(s) = \prod_{j=1}^{n} f^*_{T_n}(s) = \prod_{j=1}^{n} e^{-s\alpha^{j-1}\mu + \frac{s^2(\alpha^{j-1}\sigma)^2}{2}} = e^{-s\mu\sum_{j=1}^{n}\alpha^{j-1} + \frac{s^2\sigma^2}{2}\sum_{j=1}^{n}(\alpha^{j-1})^2}$$

$$= e^{-s\mu\left(\frac{1-\alpha^n}{1-\alpha}\right) + \frac{s^2\sigma^2}{2}\left(\frac{1-\alpha^{2n}}{1-\alpha^2}\right)}$$

and we can conclude that the distributions on the S_n are normal with mean $1-\alpha^n/1-\alpha\mu$ and standard deviation $(1-\alpha^{2n}/1-\alpha^2)^{1/2}\sigma$. Then, we observe that as n increases, both the mean and the standard deviations values converge — often quite rapidly. In addition, for any time interval, t, the sum of the quantities $F_{S_n}(t)$ will also converge. Algebraically, it is difficult to identify the convergent form, but we can easily compute approximate values of the limit numerically. As a general observation, the closer the value of α is to 1.0, the more quickly the values $F_{S_n}(t)$ decline, and therefore, the more quickly the numerical convergence of the sum.

Suppose the underlying life distribution has μ = 4000.0 and σ = 1000.0. The corresponding (approximate) limits for several values of α are shown in Table 10.5. As the value of α is decreased, the number of terms that must be computed increases. For example, for α = 0.95 and a time of 5000 hours, $M(5000)$ is determined using only two terms, $F_{S_n}(t)$, and $M(10000)$ requires six terms. On the other hand, for α = 0.75, computing $M(5000)$ requires six terms and computing $M(10000)$ requires 200 terms.

10.4 CONCLUSION

As in the previous chapter, the models presented above serve to highlight the questions one should consider in the study of repairable systems. The models in this chapter focus on the more realistic representation of the postrepair equipment state. By treating repair as instantaneous, we obtain the simplest model forms possible and develop the methods best suited to their analysis. With these methods now defined, we are ready to move on to the investigation of equipment performance when repair is noninstantaneous.

10.5 EXERCISES

1. Assume the operation of a device is to be represented using the minimal repair model with a Weibull life distribution having $\beta = 1.5$ and $\theta = 5000.0$ hours. Compute $\Pr[T_4 - T_3 > 1200]$, $\Pr[N_t = 4]$, $\Pr[N_{4000} - N_{1000} > 2]$, $\Pr[N_{6000} - N_{3000} > 2]$, and $\Pr[N_{8000} - N_{5000} > 2]$.
2. Plot the cumulative intensity function for the minimal repair model for the case in which the underlying life distribution is Gamma with $\beta = 2$ and $\lambda = 0.20$.
3. Compute and plot the distribution on the time between device renewals for the imperfect repair model having a Weibull life distribution with $\beta = 1.5$ and $\theta = 5000.0$ hours.
4. Replicate the bounds on $E[S_n]$ for the Kijima I stated in Table 10.3 by computing the values of $E[S_n, p]$ using Expression 10.32 for the case in which $\pi = 0.90$.
5. Construct and run a simulation analysis of the Kijima I model for the Weibull life distribution with $\beta = 2.75$ and $\theta = 4000.0$ hours for the case in which $\pi = 0.90$.
6. For the quasi-renewal model, assume $F_X(t)$ is Gamma with $\beta = 2.0$ and $\lambda = 0.05$, and construct the general expressions for $f_{T_n}(t)$, $F_{T_n}(t)$, and $F_{S_n}(t)$. Then construct a numerical routine to compute $M(t)$ for $t = 2000$, $t = 5000$, and $t = 8000$ when $\alpha = 0.90$.

11

Availability Analysis

The models and results described in Chapters 9 and 10 are constructed without considering the duration of the repair activity. In many situations, the focus of our analysis is upon questions for which the answers do not depend upon the duration of the repair process. This is especially true when the time a system spends down is relatively less unimportant than the fact that a failure has occurred and when the duration of repair is small or negligible in comparison to the device life length. In contrast to such devices, there are components and systems for which the duration of the repair activity has an impact on the meaningful device performance measures. For items of this sort, we must include repair times in our models.

Naturally, when we include repair time in our models, we may represent the possible repair durations in whatever manner seems most representative of actual experience. In some cases, repair time is taken to be a constant, while in most cases, repair time is treated as a random variable, and a specific distribution is selected to portray the dispersion in repair times. We will consider both of these possibilities.

We may consider that a typical sample path for a device is one in which periods of operation are terminated by device failure and that, therefore, a repair period follows each failure. Upon completion of the repair, the device is placed in

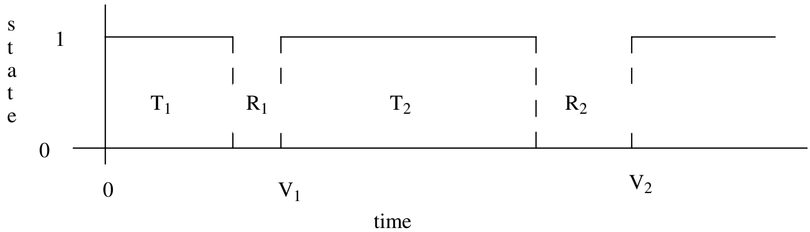


Figure 11.1 Representative sample path.

operation again. A representative sample path is shown in Figure 11.1. Note that we will modify the labels soon, but for now, we indicate that the periods of repair have durations R_j, while the operating periods are labeled as T_j. Observe also that the device's state is denoted by X_i, and that the value of the state variable is shown to be 1 when the device is operating and 0 when it is being repaired. Note further that each pair of periods, operating and repair, is shown to have total duration V_j where

$$V_j = T_j + R_j \tag{11.1}$$

It is reasonable to consider that each interval that includes an operating period and a repair period is a "cycle."

In the particular case, in which all of the operating times T_j are random variables with a common probability distribution and all of the repair times R_j are random variables with a common probability distribution, the series V_j forms a renewal process. This is because the intervals of operation T_j form a renewal process, as do the intervals of repair R_j, and thus the V_j form an alternating renewal process that is also a renewal process. For the majority of the current chapter, the cycles are assumed to form a renewal process, and the types of analysis that can be performed are discussed. Toward the end of the next chapter, we examine the relatively few results that have been constructed for the nonrenewal case.

To study the cycles in the renewal case, we first modify the definition of the quantities V_j. In general, our analyses

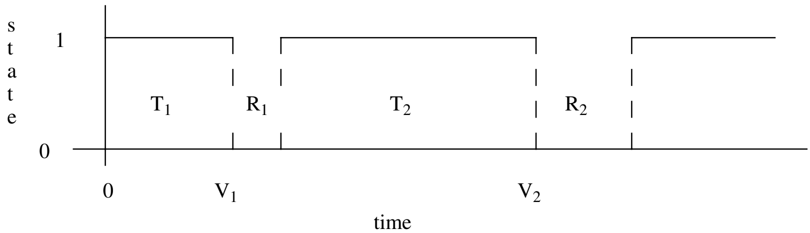


Figure 11.2 Representative sample path with revised labeling of the V_j.

are based on an interest in the times of device failure rather than in times of device restart. Therefore, we modify the labels on the sample paths so that the V_j correspond to failure times. This is illustrated in Figure 11.2. The difference is quite subtle. The change is that the values of the V_j now correspond to failure times. That is,

$$\begin{aligned} V_1 &= T_1 \\ V_2 &= R_1 + T_2 \\ &\vdots \\ V_j &= R_{j-1} + T_j \\ &\vdots \end{aligned} \tag{11.2}$$

Using this definition, the series V_j forms a "modified" renewal process because the distributions for all of the V_j other than V_1 are the same, and that for V_1 is different.

As a matter of convention, let the distributions $F_{T_j}(t)$ represent the distributions on the lengths of the operating intervals (which could be just the life distribution), and let $G_{R_j}(t)$ represent the distributions on the lengths of the repair intervals. In that case, the distributions on the durations of the cycles are constructed as the convolution of $F_{T_j}(t)$ and $G_{R_j}(t)$ and are denoted by $H_{V_j}(t)$. Thus, in general,

$$H_{V_j}(t) = F_{T_j}(t) * G_{R_{j-1}}(t) = \int_0^t G_{R_{j-1}}(t-u)dF_{T_j}(u) \tag{11.3}$$

When the distributions are common for all cycles, the subscript "j" is dropped.

An interesting consequence of the relabeling of the sample path is that the renewal functions for the expected number of failures and the expected number of repairs are different. We will exploit that difference in our analysis later. For now, note that the Laplace transform for the expected number of repair completions during and interval $(0, t)$ is given by

$$M^*_{G_R}(s) = \frac{f^*_T(s)g^*_R(s)}{s\left(1 - f^*_T(s)g^*_R(s)\right)} \tag{11.4}$$

while the transform for the expected number of failures during an interval $(0, t)$ is

$$M^*_{H_V}(s) = \frac{f^*_T(s)}{s\left(1 - f^*_T(s)g^*_R(s)\right)} \tag{11.5}$$

The expected number of failures and the expected number of points V_j are the same. These arise from the applicable forms of the key renewal theorem, which are

$$M_{G_R}(t) = H_V(t) + \int_0^t M_{G_R}(t-u)h_V(u)du \tag{11.6}$$

for the expected number of repair completions and:

$$M_{H_V}(t) = F_T(t) + \int_0^t M_{H_V}(t-u)h_V(u)du \tag{11.7}$$

for the expected number of failures.

11.1 AVAILABILITY MEASURES

Regardless of whether or not the cycles form a renewal process, the fact that there are periods during which the device is not functioning leads naturally to the definition of "availability" as a measure of device performance. In fact, four distinct but related availability measures have been defined. The basic one is

Defn. 11.1: The (*point*) *availability* at time t, $A(t)$, for a device is the probability that it is functioning ($X(t) = 1$) at the time. Thus,

$$A(t) = \Pr[X(t) = 1] = E[X(t)] \tag{11.8}$$

The distinction that this is the point availability is important. Nevertheless, $A(t)$ is usually called simply the availability, and this is done here unless doing so is ambiguous. The other three availability measures are:

Defn. 11.2: The *limiting availability* for a device is the probability is the limit of the point availability function. That is,

$$A_\infty(t) = \lim_{t\to\infty} A(t) \tag{11.9}$$

The limiting availability can be very useful. As we will see, many devices experience an interval of transience before they "settle down" into a consistent pattern of operation. At that point, the devices often display availability behavior that is stable and similar (or equal) to the limiting form. In addition, there are many analytical cases in which the point availability is very difficult (or impossible) to compute, but the limiting availability measure is manageable.

It may also be useful at certain times to compute averages so we have:

Defn. 11.3: The *average availability* over an interval (t_1, t_2) for a device is

$$\bar{A}(t_1, t_2) = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} A(u)du \tag{11.10}$$

The average availability is often evaluated for $t_1 = 0$. In addition, we have

Defn. 11.4: The *limiting average availability* for a device is the limit of the average availability over (0, t). That is,

$$\bar{A}_\infty = \lim_{t\to\infty} \bar{A}(0, t) \tag{11.11}$$

Each of the four measures has utility in specific cases. In general, the point availability is the most informative measure, but it is usually the most difficult to obtain. The most commonly used of the measures is the limiting availability. The primary reason for its popularity is that, in the case in which the operating and repair cycles form a renewal process, it is easily computed as

$$A_\infty = \frac{E_F[T]}{E_F[T] + E_G[T]} = \frac{MTTF}{MTTF + MTTR} \tag{11.12}$$

where *MTTF* is the "mean time to failure," and *MTTR* is the "mean time to repair." Of course, when no repair is possible, the availability measures reduce to their corresponding reliability terms.

For most applications, we should find that $\bar{A}_\infty = A_\infty$. Also, $\bar{A}(t_1, t_2)$ is the proportion of the time the device operates during the interval (t_1, t_2). A further point that is somewhat subtle is the fact that, for a population of devices, each of the availability measures is an expected value relative to the number (or proportion) of the members of the population that are operational. The distribution on the frequency for which the availability measures are expected values is the binomial [57].

Given the elementary definitions stated above, we may proceed to construct an availability function. The analysis of availability can become quite intricate, so it is usually performed relative to a very carefully defined statement of the assumed operating and repair scenario. This is because very minor differences in operation plan, and particularly in the assumed failure and repair distributions, may have quite pronounced effect on the identity of the appropriate model and its solution.

For now, we assume that each repair operation returns a device to a "good as new" state, so that all of the operating periods have the same distribution on duration. Similarly, we assume that all of the repair activities have a common distribution on duration. Therefore, as noted above, each cycle has the same probability distribution on its length, and the cycles form a renewal process. Therefore, the availability function may be defined as the sum of two probabilities:

1. The probability that the device has never failed and is thus still functioning
2. The probability that a new cycle was started at some recent point in time and no failure has occurred since then

Algebraically, this is

$$A(t) = \bar{F}_T(t) + \int_0^t \bar{F}_T(t-u)m_{H_V}(u)du \tag{11.13}$$

where, as indicated above, $H_V(t)$ is the convolution of the distributions on operating intervals and on repair times. Note particularly the similarity of the availability function to the key renewal theorem.

Generally, when we can, we use Laplace transforms to obtain the final form of the availability function. In that case, we construct

$$\begin{aligned} A^*(s) &= \frac{1}{s} - F_T^*(s) + \left(\frac{1}{s} - F_T^*(s)\right)m_{H_V}^*(s) \\ &= \frac{1}{s}\left(1 - f_T^*(s)\right)\left(1 + m_{H_V}^*(s)\right) \end{aligned} \tag{11.14}$$

and the inverse transform gives the availability function. Note that Equation 9.16 of Chapter 9 implies that we may use

$$A^*(s) = \frac{1}{s}\left(1 - f_T^*(s)\right)\frac{1}{1 - f_T^*(s)g_R^*(s)} \tag{11.15}$$

instead of Expression 11.14, especially if it is easier to avoid determining the specific form of the renewal density. In addition, as indicated previously, we can exploit the difference in the repair completion and the failure time renewal functions to obtain an additive form. Specifically,

$$A^*(s) = M_{G_R}^*(s) - M_{H_V}^*(s) + \frac{1}{s} \tag{11.16}$$

The advantage of this form is that its inverse is

$$A(t) = M_{G_R}(t) - M_{H_V}(t) + 1 \tag{11.17}$$

Also, we again avoid specifically constructing the renewal density.

11.2 EXAMPLE COMPUTATIONS

Nearly all of the possible realizations of the availability model are algebraically and computationally quite difficult. The exception is the case in which both the life distribution and the repair time distribution are exponential and the repairs return the system to a "good as new" state. In order to demonstrate the method of analysis, the availability function for this renewal case is constructed here. Following that analysis, a more difficult case is examined.

11.2.1 The Exponential Case

To start, assume the exponential life distribution has parameter λ_f, and the exponential repair time distribution has parameter λ_r. It is worthwhile to construct the distribution on cycle length, as this illustrates the process of constructing convolutions. As we know:

$$h_V^*(s) = f_T^*(s) g_R^*(s) = \frac{\lambda_f}{s+\lambda_f}\frac{\lambda_r}{s+\lambda_r} = \frac{\lambda_f \lambda_r}{s^2 + (\lambda_f + \lambda_r)s + \lambda_f \lambda_r}$$

For this transform, we use the method of partial fractions to solve

$$\frac{1}{s^2 + (\lambda_f + \lambda_r)s + \lambda_f \lambda_r} = \frac{A}{s+\lambda_f} + \frac{B}{s+\lambda_r}$$

for the forms

$$A = \frac{1}{\lambda_r - \lambda_f}$$

and

$$B = \frac{-1}{\lambda_r - \lambda_f}$$

Thus, the inverse of the transform is

$$h_V(t) = \frac{\lambda_f \lambda_r}{\lambda_r - \lambda_f}\left(e^{-\lambda_f t} - e^{-\lambda_r t}\right)$$

with the corresponding distribution function

$$H_V(t) = 1 - \frac{\left(\lambda_r e^{-\lambda_f t} - \lambda_f e^{-\lambda_r t}\right)}{\lambda_r - \lambda_f}$$

Observe that this is neither an exponential nor a gamma distribution, but that it is a proper distribution function, and it provides an unambiguous model of the dispersion in the lengths of the cycles of operation and repair. Naturally, it also correctly yields

$$E_{H_V}[T] = E_F[T] + E_R[T]$$

Next, in order to construct the availability function, we use Expression 11.15:

$$A^*(s) = \frac{1}{s}\left(1 - f_T^*(s)\right)\frac{1}{1 - f_T^*(s) g_R^*(s)}$$

$$= \frac{1}{s}\left(1 - \frac{\lambda_f}{s + \lambda_f}\right)\left(1 - \frac{\lambda_f}{s + \lambda_f}\frac{\lambda_r}{s + \lambda_r}\right)^{-1}$$

which we can reduce to

$$A^*(s) = \frac{s + \lambda_r}{s(s + \lambda_f + \lambda_r)}$$

Using the partial fraction expansion:

$$\frac{s+\lambda_r}{s(s+\lambda_f+\lambda_r)} = \frac{A}{s} + \frac{B}{(s+\lambda_f+\lambda_r)}$$

we obtain

$$A = \frac{\lambda_r}{\lambda_r+\lambda_f}$$

and

$$B = \frac{\lambda_f}{\lambda_r+\lambda_f}$$

and the inverse transform is

$$A(t) = \frac{\lambda_r}{\lambda_f+\lambda_r} + \frac{\lambda_f}{\lambda_f+\lambda_r} e^{-(\lambda_f+\lambda_r)t}$$

This result is quite useful for examining the behavior of the availability measures for the renewal case. Note first that the availability function is comprised of a constant term and a term that diminishes over time, so the function displays an initial transience and then settles down to a stable and essentially constant value. This is illustrated in Figure 11.3 for $\lambda_f = 0.01$ and $\lambda_r = 0.1$. The limiting value is

$$A_\infty = \frac{\lambda_r}{\lambda_f+\lambda_r} = 0.909$$

Note also that we may use Expression 11.5 to compute

$$\bar{A}(t_1, t_2) = \frac{\lambda_r}{\lambda_f+\lambda_r} + \frac{\lambda_f}{(t_2-t_1)(\lambda_f+\lambda_r)^2}\left(e^{-(\lambda_f+\lambda_r)t_1} - e^{-(\lambda_f+\lambda_r)t_2}\right)$$

so, for example, $\bar{A}(10, 20) = 0.927$. Also, we may note that $\bar{A}_\infty = A_\infty$.

Examine the limiting availability that we obtained for the exponential case. Specifically, we can restate the limiting availability as

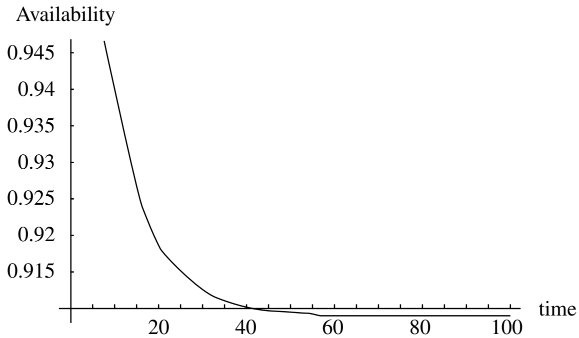


Figure 11.3 Example availability function.

$$A_\infty = \frac{\lambda_r}{\lambda_f + \lambda_r} = \frac{1/E_G[T]}{1/E_F[T] + 1/E_G[T]} = \frac{E_F[T]}{E_F[T] + E_G[T]}$$

which is the form given in Expression 11.12.

As we have seen above, the exponential model provides a relatively simple illustration of all of the basic availability analyses. It must be emphasized that other choices of life and repair time distributions can be analyzed, but the other cases often require considerable computational effort.

On the other hand, there are quite a number of nonexponential cases in which the bounds described in Chapter 9 can be used to provide information on device performance. For cases in which the life distribution and the repair time distribution are both IFR, the distribution on cycle length, $H_V(t)$, will also be IFR. In that case, we can compute bounds on the lengths of cycles and on the number of cycles, using Expressions 9.25 through 9.29 and Expression 9.32.

11.2.2 A Numerical Case

If the distribution that best represents the life lengths of a device is not exponential, the computation of the point availability function can be quite taxing. However, given modern computing power, the calculations are frequently manageable.

The advantages of the capabilities of modern computers and software should not be taken lightly, as many of the interesting results we can now obtain were completely impossible in the past. In any case, the computation of the point availability function may be pursued in any of several ways. One approach is to construct the relevant Laplace transforms and then use numerical algorithms to invert $A^*(s)$. An alternate approach is to directly perform the integration of Expression 11.13 numerically.

Consider an example. Suppose the life lengths for a population of a device are well represented by a Weibull distribution having parameters β = 2.0 and θ = 200 hours. Suppose further that repair of the device involves replacement and takes a random time, for which an exponential distribution with λ = 0.10 provides a reasonable model. For these distributions, we first note that

$$A_\infty = \frac{E_F[T]}{E_F[T]+E_G[T]} = \frac{177.245}{177.245+10.0} = 0.94659$$

Next, we can evaluate the point availability by performing the numerical integration of Expression 11.13, in which we make the substitution

$$m_{H_V}(t) = \sum_{k=1}^{\infty} \int_0^t f_F^{(k)}(u) g_R^{(k)}(t-u)\,du \tag{11.18}$$

and then replace $f_T^{(n)}(t)$ with its approximate form as defined in Appendix B. The resulting values of the function are fully accurate until we introduce the numerical error associated with the integration and the error resulting from the finite truncation of the infinite sums. The actual numerical effort involved in this analysis is quite taxing. However, it is manageable. Several of the values of the point availability obtained for the given distributions are shown in Table 11.1. The values show a transient interval and a relatively rapid convergence to the limiting value. In view of the fact that both the Weibull life distribution and the exponential repair time

Table 11.1 Computed Values of the Availability Function for Weibull Failures and Exponential Repairs

Time	$A(T)$
50	0.9807
100	0.9614
150	0.9498
200	0.9454
250	0.9451
300	0.9466
350	0.9466
400	0.9466

distribution center around exponential terms, the rapid convergence is exactly the type of behavior one expects to see.

11.3 SYSTEM-LEVEL AVAILABILITY

Naturally, it is most often the case that we are interested in the availability of a system for which repair involves "servicing" one or several components of a multicomponent system. Examples of systems for which availability is important and repair frequently involves component or module replacement include aircraft, computers, production equipment, and motorized vehicles. For the present analysis, consider that the repair activity involves the replacement of a component. Of course, there are very many possible operating scenarios, each of which may involve distinct models. Consider two cases here.

For the first case, suppose that a system is comprised of independent components arranged in an arbitrary configuration and that each component has an associated life distribution and repair time distribution. Also assume that components are replaced on failure, but that while any single component is being repaired, the other components continue to operate and thus to age. For the second case, we will assume a series system in which the functioning components do not age during the replacement of a failed component.

The first case fits very well with the reliability and system status construction we studied earlier. Specifically, suppose that, for component j, the life distribution is $F_{j,T}(t)$ and the repair time distribution is $G_{j,R}(t)$, so that the renewal function for that component is $M_{H_{j,V}}(t)$. Then the availability function for the component is $A_j(t)$ where, as before,

$$A_j(t) = \Pr[X_j(t) = 1]$$

The availability function $A_j(t)$ is computed as the solution to Expression 11.13 without regard to the other components. Then, since the system status is

$$A_S(t) = E[\phi(\underline{X}(t))] = \phi(E[\underline{X}(t)]) \tag{11.19}$$

the systems availability functions is

$$A_S(t) = \phi(A_1(t), A_2(t)..., A_n(t)) \tag{11.20}$$

Thus, we compute the system availability in exactly the same manner as our calculation of system reliability. All of the computation rules we developed in Chapter 3, including the rules for calculating bounds for general structures, apply. In addition, these rules apply to the limiting availabilities as well as to the point availabilities.

Consider two examples. Suppose that each of the components of a five-component series structure has an exponential life distribution with the parameters λ_{f1} = 0.833, λ_{f2} = 1.250, λ_{f3} = 2.500, λ_{f4} = 1.111, and λ_{f5} = 0.200, and that for each component, the distribution on repair time is exponential with the parameters λ_{r1} = 0.0625, λ_{r2} = 0.040, λ_{r3} = 0.100, λ_{r4} = 0.025, and λ_{r5} = 0.050. Then, we know from our earlier analysis that

$$A_j(t) = \frac{\lambda_{jr}}{\lambda_{jf} + \lambda_{jr}} + \frac{\lambda_{jf}}{\lambda_{jf} + \lambda_{jr}} e^{-(\lambda_{jf} + \lambda_{jr})t}$$

and

$$A_{j\infty} = \frac{\lambda_{jr}}{\lambda_{jf} + \lambda_{jr}}$$

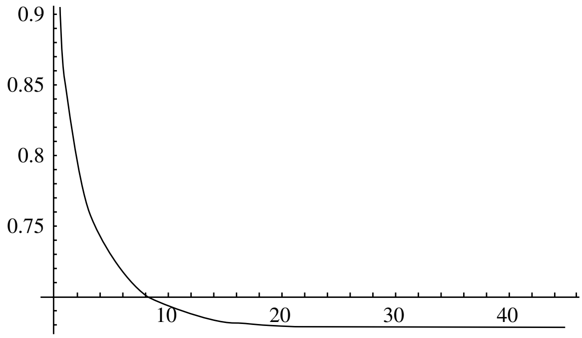


Figure 11.4 Point availability function for an example series system.

Therefore, since the system is a series system,

$$A_{S\infty} = \prod_{j=1}^{5} A_{j\infty} = \prod_{j=1}^{5} \frac{\lambda_{jr}}{\lambda_{jf} + \lambda_{jr}} =$$

$$= (0.930)(0.969)(0.962)(0.998)(0.800) = 0.678$$

and the point availability function is the one shown in Figure 11.4. We can see that the product form of the structure function implies a fairly rapid decline in the value of the exponential term, and the system thus "settles down" into its limiting behavior in a short time.

As an example of the more complicated analysis, we might encounter, consider the bridge structure of Chapter 2 and assume each of the components has a Weibull life distribution with parameters β = 2.0 and θ = 200 hours and an exponential repair time distribution with parameter λ = 0.10. As we have already computed the availability for a single device with these characteristics, we know that $A_j(50) = 0.9807$ and $A_j(200) = 0.9454$. We may also recall that the series-parallel bounds, the min path upper-min cut lower bounds, and the mini-max bounds on reliability for the bridge structure with identical components are

$$r^5 \le R_s \le 1-(1-r)^5$$

$$\left(1-(1-r)^2\right)^2\left(1-(1-r)^3\right)^2 \le R_s \le \left(1-(1-r^2)^2(1-r^3)^2\right)$$

and

$$r^2 \le R_s \le 1-(1-r)^2$$

respectively. Replacing the reliabilities with availability terms yields

$$\left(A_j(t)\right)^5 \le A_s(t) \le 1-\left(1-A_j(t)\right)^5$$

$$\left(1-\left(1-A_j(t)\right)^2\right)^2\left(1-\left(1-A_j(t)\right)^3\right)^2 \le A_s(t)$$

$$\le \left(1-\left(1-\left(A_j(t)\right)^2\right)^2\left(1-\left(A_j(t)\right)^3\right)^2\right)$$

and

$$\left(A_j(t)\right)^2 \le A_s(t) \le 1-\left(1-A_j(t)\right)^2$$

Therefore, we can compute the availability bounds:

$$0.9072 \le A_s(50) \le 1.0$$

$$0.9992 \le A_s(50) \le 0.9999$$

$$0.9618 \le A_s(50) \le 0.9996$$

and

$$0.7552 \le A_s(200) \le 1.0$$

$$0.9937 \le A_s(200) \le 0.9997$$

$$0.8938 \le A_s(200) \le 0.9970$$

Of course, we also have the limiting availabilities of

$$A_{j\infty} = \frac{E[T_j]}{E[R_j] + E[T_j]} = \frac{177.245}{10.0 + 177.245} = 0.947$$

so we can compute the bounds:

$$0.760 \le A_{s\infty} \le 1$$

$$0.994 \le A_{s\infty} \le 0.969$$

and

$$0.896 \le A_{s\infty} \le 0.997$$

Thus, we can see that, for the assumed operating scenario, the results we developed earlier using the system structure can be very useful and informative.

A second model for which we can obtain results is that in which we have an m-component series system. We assume that the failure of any component stops the operation of the system and, thus, of the other components. The other components do not age while the failed component is being replaced. Thus, when system operation is resumed, the replaced component is new, and the other components have the age they had when the replaced component failed. Several key structural results for this model were developed by Barlow and Proschan [11]. These results are general, in that they apply to any choice of continuous life and repair time distributions.

The sample path shown in Figure 11.1 at the start of this chapter is representative of the operating experience for a series system. The individual failure and repair times may (usually do) correspond to the failures and replacements of different components. Thus, let T_{ji} represent the life length of the i^{th} copy of component j used in the system, and let R_{ji} represent the i^{th} replacement time for the j^{th} component. Next, let $U(t)$ represent the system operating time (up time) during the real time interval $(0, t)$. Note that $U(t)$ is not an availability measure, but that for any realization of the system sample path, $U(t)/t$ is the proportion of the time the system functioned and, hence, is the average availability for that sample path. Now, Barlow and Proschan [11] prove that

$$\bar{A}_{S\infty} = \lim_{t\to\infty} \frac{U(t)}{t} = \left(1 + \sum_{j=1}^{m} \frac{E_{G_j}[R]}{E_{F_j}[T]}\right)^{-1} \quad (11.21)$$

That is, we take the ratio of the expected repair time to the expected life length for each component and sum across components. Add this to one and take the reciprocal, and the result is the limiting average system availability. In addition, for typical choices of the life distributions — those that are continuous — the limiting system availability is the same as the limiting average system availability. Also, the limiting values for the component replacement rates are the respective reciprocals of the mean life times

$$\lim_{t\to\infty} \left\{ E[{}^{N_j(U(t))}\!/_{U(t)}] \right\} = \frac{1}{E_{F_j}[T]} \quad (11.22)$$

where, as in our previous analyses,

$$S_{jn} = \sum_{i=1}^{n} T_{ji}$$

and

$$N_j(t) = \sup\{n \mid S_{jn} \le t\}$$

A conceptually equivalent statement to Expression 11.22 is that the average life length of a component divided by the total up time equals (in the limit) the average number of copies of the component that are used. Additional useful results are

$$U(t) \approx tA_{S\infty} \quad (11.23)$$

and

$$\lim_{t\to\infty} \left\{ \frac{N_j(t)}{t} \right\} \approx \frac{\bar{A}_{S\infty}}{E_{F_j}[T]} \quad (11.24)$$

Over time, the up time is approximated by the product of time and limiting average system availability, and the number of replacements for any component is approximately the limiting average availability divided by the average lift length. (Note that dividing both sides of Expression 11.24 gives (approximately) Expression 11.22.)

At the system level, we can state that the average system up time per cycle converges in the limit to

$$\bar{U} = \left(\sum_{j=1}^{m} 1 \Big/ E_{F_j}[T] \right)^{-1} \tag{11.25}$$

and the corresponding limit for the average time down per cycle is

$$\bar{D} = \bar{U} \left(\sum_{j=1}^{m} \frac{E_{G_j}[R]}{E_{F_j}[T]} \right) \tag{11.26}$$

Clearly, these imply the average cycle length, and they are consistent with the average availability expressions.

Consider again the example of a system comprised of $m = 5$ components in series, each of which has an exponential life distribution with the parameters $\lambda_{f1} = 1.20$, $\lambda_{f2} = 0.800$, $\lambda_{f3} = 0.400$, $\lambda_{f4} = 0.90$, and $\lambda_{f5} = 5.00$ and an exponential repair time distribution with parameters $\lambda_{r1} = 19.01$, $\lambda_{r2} = 25.00$, $\lambda_{r3} = 10.00$, $\lambda_{r4} = 40.00$, and $\lambda_{r5} = 20.00$. In this case, we find

$$\bar{A}_{S\infty} = \left(1 + \frac{(0.0625)}{(0.833)} + \frac{(0.0640)}{(1.250)} + \frac{(0.100)}{(2.500)} + \frac{(0.025)}{(1.111)} + \frac{(0.050)}{(0.200)} \right)^{-1}$$

$$= 0.695$$

$$\bar{U} = \left(\sum_{j=1}^{m} 1 \Big/ E_{F_j}[T] \right)^{-1} = (0.833 + 1.25 + 2.50 + 1.111 + 0.20)^{-1}$$

$$= 0.1668$$

$$\bar{D} = \bar{U}(0.4387) = 0.0732$$

so the average cycle length is 0.24 time units.

There are two interesting final observations concerning the system availability measures. First, Expressions 11.21, 11.25, and 11.26 jointly imply that

$$A_{S\infty} = \frac{\bar{U}}{\bar{U} + \bar{D}}$$

as we would expect to be the case. Similarly, if the number of components of the system is $m = 1$, the limiting average system availability of Expression 11.21 reduces to

$$\bar{A}_{S\infty} = \left(1 + \frac{E_G[R]}{E_F[T]}\right)^{-1} = \frac{E_F[T]}{E_F[T] + E_G[R]}$$

and this is also as we would expect.

11.4 THE NONRENEWAL CASES

Very few availability results have been developed for the cases in which repair does not imply renewal. Numerous authors have used cost models rather than availability models to describe the consequences of downtime. The reason for the shift to cost models is the complexity (or impossibility) of the probability analysis for availability. Some of these cost models are discussed in the next chapter. In the meantime, we can construct some single unit availability results for the imperfect repair model.

Recall that under an imperfect repair regime, following repair, a device is returned to a good as new condition by a perfect repair with probability p and remains in a bad as old condition with probability $q = 1 - p$. The generalization of this model, defined by Block, Borges, and Savits [52] and discussed in Chapter 10, is to make the probability of perfect repair age dependent. Then, the probability of perfect repair is $p(t)$. As discussed in Chapter 10, the distribution on the total operating time from a renewal point until a failure that is followed by a perfect repair for the first time is (Expression 10.10):

$$F_p(t) = 1 - \left(\bar{F}_T(t)\right)^p = 1 - e^{-pZ_T(t)} \tag{11.27}$$

This expression applies when the minimal repair actions are instantaneous. However, it also represents the distribution on total operating time for the case in which minimal repairs have nonzero durations. In addition, it is useful to note that the average time of operation per renewal cycle may be determined as

$$E_p[T] = \int_0^\infty \bar{F}_p(u)du \tag{11.28}$$

Now, if the minimal repair times have distribution $G_m(t)$, and if it is the n^{th} device failure that produces the first perfect repair, then the total time spent in minimal repair will have the distribution $G_m^{(n-1)}(t)$, the $n - 1$ fold convolution of $G_m(t)$. For this conceptual model, Iyer [58] has shown that the distribution on the total time devoted to minimal repairs in any renewal cycle is

$$G(t) = e^{-\Lambda_m(t)} \sum_{n=1}^{\infty} \frac{\left(\Lambda_m(t)\right)^{n-1}}{(n-1)!} G_m^{(n-1)}(t) \tag{11.29}$$

where

$$\Lambda_m(s) = \int_0^s q(u) z_T(u)du \tag{11.30}$$

represents the cumulative intensity function for the occurrence of a failure with a minimal repair response. Clearly, the distribution in Expression 11.29 is constructed by using a Poisson distribution on the number of minimal repair events as the mixing distribution for the extent of the minimal repair time convolution.

The convolution of the distribution $G(t)$ with the distribution on total operating time per cycle yields the time until the occurrence of a failure that is to be followed by a perfect repair. That is, the distribution on time until a perfect repair is started

$$F_{p*}(t) = \int_0^t G(t-u) f_p(u) du \tag{11.31}$$

Similarly, the distribution on the total duration of the renewal cycle is the convolution of this distribution with the one on perfect repair time, say, $G_p(t)$.

We may also determine the limiting device (average) availability using the expectations for the durations already defined. To do this, note first that, in general, each renewal cycle will include say N periods of device operation terminated by a device failure. The first $N - 1$ of the operating periods are followed by periods of minimal repair, and the last failure is followed by a period of perfect repair. Now Iyer [58] has shown that

$$E[N] = 1 + \int_0^\infty \Lambda_m(u) \bar{F}_p(u) du \tag{11.32}$$

and, since $E[N-1] = E[N] - 1$, the limiting availability is

$$A_\infty = \frac{E_{F_p}[T]}{E_{F_p}[T] + E[N-1]E_{G_m}[T] + E_{G_p}[T]} \tag{11.33}$$

That is, the average up time divided by the average cycle length gives us the average availability. The average cycle length is comprised of the average up time, plus the average number of minimal repair intervals times their average duration, plus the average duration of the perfect repair interval.

Notice that we can also state that the average fraction of the time the device is undergoing minimal repair is

$$A_\infty = \frac{E[N-1]E_{G_m}[T]}{E_{F_p}[T] + E[N-1]E_{G_m}[T] + E_{G_p}[T]}$$

and the average fraction of the time the device is undergoing perfect repair is

$$A_\infty = \frac{E_{G_p}[T]}{E_{F_p}[T] + E[N-1]E_{G_m}[T] + E_{G_p}[T]}$$

Finally, we note that, for the simpler case in which $p(t) = p$, the probability of perfect repair is age independent, so the expected number of failures to a perfect repair is geometric, and

$$E[N] = 1/p \tag{11.34}$$

In addition, the distribution $G(t)$ reduces to

$$G(t) = \sum_{n=1}^{\infty} \frac{(1-q)q^{n-1}}{(1-q^n)} G_m^{(n-1)}(t) \tag{11.35}$$

Other simplifications are also consistent with our results. If we take $p = 0$, we obtain the minimal repair model either with or without instantaneous minimal repair times. If we take $p = 1$, we have the renewal case. Assuming the basic distributions are exponential yields several simple models that all behave as Markov processes, as we would expect.

11.5 MARKOV MODELS

To conclude this chapter, note that there is an entirely different approach than the ones discussed above to modeling the time evolution of repairable systems. Many people, and most notably Birolini [59], have used Markov chains and Markov processes to represent device- and system-level behaviors and have used the Markov models to obtain availability or cost measures. The core idea for the Markov models is to represent system status in a manner like the one shown in Figure 11.5. In the figure, the failure hazard is represented by $\lambda(t)$, and the repair hazard is denoted by $\mu(t)$. These functions represent the intensity of transition to and from the functioning state (state 1). Under this format, the device operation is represented as a continuous time Markov process with the Chapman-Kolmogorov forward differential equations [60] being

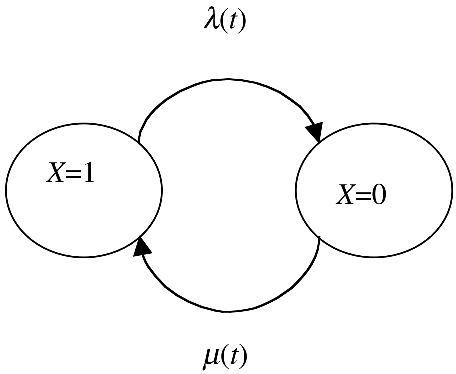


Figure 11.5 Two-state transition diagram.

$$\begin{aligned}
\frac{d}{dt}p_{1,1}(t) &= -\lambda(t)p_{1,1}(t) + \mu(t)p_{2,1}(t) \\
\frac{d}{dt}p_{1,2}(t) &= -\mu(t)p_{1,2}(t) + \lambda(t)p_{1,1}(t) \\
\frac{d}{dt}p_{2,1}(t) &= -\lambda(t)p_{2,1}(t) + \mu(t)p_{2,2}(t) \\
\frac{d}{dt}p_{2,2}(t) &= -\mu(t)p_{2,2}(t) + \lambda(t)p_{2,1}(t)
\end{aligned} \tag{11.36}$$

where $p_{i,j}(t)$ is the probability that the process passes from state i to state j during an interval of length t. For the case in which the intensity functions are constant, the fact that $p_{1,1}(t)+(t)p_{1,2}(t)=1$ implies that the first of the equations (11.36) becomes

$$\frac{d}{dt}p_{1,1}(t) = -(\lambda+\mu)p_{1,1}(t)+\mu$$

for which the Laplace transform is

$$p^*_{1,1}(s) - p_{1,1}(0) = -(\lambda+\mu)p^*_{1,1}(s)+\mu s$$

so

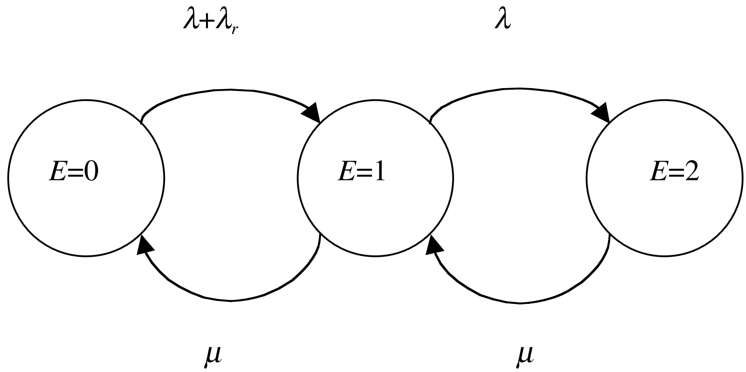


Figure 11.6 State transition diagram for the repairable parallel system.

$$p^*_{1,1}(s) = \frac{s+\mu}{s(s+\lambda+\mu)}$$

which is the form given in Section 11.2.1. Clearly, the result may be constructed using either approach.

For some of the more complicated cases, the use of the Markov process model is easier than the direct analysis of the renewal process model. As a case in point, consider a repairable parallel system of two components. Assume that two identical components are arranged in a parallel configuration, and that the system is operated so that one of the components is functioning and has constant failure hazard λ, while the second component is in reserve and is therefore subject to a constant but reduced failure hazard, say λ_r. Suppose that, whenever a component fails, repair is undertaken immediately. The repair time is exponential with intensity μ, and if the repair is completed before a second component failure, the system continues to function with the repaired component in the reserve role. System failure occurs when a component failure precedes the completion of a component repair. This seemingly simple system is actually quite difficult to analyze. The state transition diagram for the representative Markov process is shown in Figure 11.6. The state space E = {0, 1, 2} represents the number of failed components. The associated Chapman-Kolmogorov forward differential equations are the following nine equations. To evaluate system availability, we need only solve the first three equations, because the availability corresponds to the time spent in states 1 and 2.

$$\begin{aligned}
\frac{d}{dt}p_{0,0}(t) &= -(\lambda+\lambda_r)p_{0,0}(t)+\mu p_{0,1}(t)\\
\frac{d}{dt}p_{0,1}(t) &= -(\lambda+\mu)p_{0,1}(t)+\mu p_{0,2}(t)+(\lambda+\lambda_r)p_{0,0}(t)\\
\frac{d}{dt}p_{0,2}(t) &= -\mu p_{0,2}(t)+\lambda p_{0,1}(t)\\
\frac{d}{dt}p_{1,0}(t) &= -(\lambda+\lambda_r)p_{1,0}(t)+\mu p_{1,1}(t)+\lambda p_{1,2}(t)\\
\frac{d}{dt}p_{1,1}(t) &= -(\lambda+\mu)p_{1,1}(t)+\mu p_{1,2}(t)+(\lambda+\lambda_r)p_{1,0}(t)\\
\frac{d}{dt}p_{1,2}(t) &= -\mu(t)p_{1,2}(t)+\lambda(t)p_{1,1}(t)\\
\frac{d}{dt}p_{2,0}(t) &= -(\lambda+\lambda_r)p_{2,0}(t)+\mu p_{2,1}(t)\\
\frac{d}{dt}p_{2,1}(t) &= -(\lambda+\mu)p_{2,1}(t)+\mu p_{2,2}(t)+(\lambda+\lambda_r)p_{2,0}(t)\\
\frac{d}{dt}p_{2,2}(t) &= -\mu p_{2,2}(t)+\lambda p_{2,1}(t)
\end{aligned} \tag{11.37}$$

Keeping in mind the fact that

$$p_{j,0}(t)+p_{j,1}(t)+p_{j,2}(t)=1$$

we use the same method as for the previous model to obtain

$$\begin{aligned}
(s+\lambda+\lambda_r)p^*_{0,0}(s)-\mu p^*_{0,1}(s) &= 1\\
-(\lambda+\lambda_r-\mu)p^*_{0,0}(s)+(s+\lambda+2\mu)p^*_{0,1}(s) &= \mu/s
\end{aligned} \tag{11.38}$$

The solution of these two equations requires considerable effort but finally one finds the solutions to be

$$p_{0,0}(t)=\frac{\left(\mu^2+\left\{\frac{(\lambda+\lambda_r)(\lambda+\mu)}{2}+\frac{(\lambda+\lambda_r)(2\lambda\mu-\lambda_r\mu-\lambda\lambda_r)}{2\sqrt{4\lambda\mu+\lambda_r^2}}\right\}\left\{e^{\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}-e^{-\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}\right\}e^{-\frac{t}{2}(2\lambda+2\mu+\lambda_r)}\right)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2}$$

$$p_{0,1}(t)=\frac{(\lambda+\lambda_r)\left(\mu+\left\{-\frac{\mu}{2}+\frac{(2\lambda^2+2\lambda\lambda_r+\lambda_r\mu)}{2\sqrt{4\lambda\mu+\lambda_r^2}}\right\}\left\{e^{\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}-e^{-\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}\right\}e^{-\frac{t}{2}(2\lambda+2\mu+\lambda_r)}\right)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2}$$

Assuming the system starts with both components being new, the sum of these two probabilities is the availability function

$$A(t)=\frac{\left(\mu(\lambda+\lambda_r+\mu)\right)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2}+\frac{\left(\left\{\frac{\lambda(\lambda+\lambda_r)}{2}+\frac{\lambda(2\lambda+2\mu+\lambda_r)(\lambda+\lambda_r)}{2\sqrt{4\lambda\mu+\lambda_r^2}}\right\}\left\{e^{\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}-e^{-\frac{t}{2}\sqrt{4\lambda\mu+\lambda_r^2}}\right\}e^{-\frac{t}{2}(2\lambda+2\mu+\lambda_r)}\right)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2} \tag{11.39}$$

and the complement is the solution for $p_{0,2}(t)$. Clearly, the limiting availability is

$$A_\infty=\frac{\left(\mu(\lambda+\lambda_r+\mu)\right)}{(\lambda+\lambda_r)(\lambda+\mu)+\mu^2} \tag{11.40}$$

Note that if we set $\lambda_r = 0$, the model and its results represent the standby redundant case in which the second component is not activated until the first one fails. For that case, the limiting availability reduces to

$$A_\infty = \frac{\left(\mu(\lambda + \mu)\right)}{\lambda(\lambda + \mu) + \mu^2}$$

Similarly, taking $\lambda_r = \lambda$, implies active redundancy, in which the second component ages while waiting to be used. In that case, the limiting availability is

$$A_\infty = \frac{\left(\mu(2\lambda + \mu)\right)}{2\lambda(\lambda + \mu) + \mu^2}$$

For both of these special cases, the corresponding forms for the point availability function apply.

The Markov process models are used widely to represent system behavior. It is appropriate to emphasize that the above case is the simplest one of its class. A more general case is to assume a k out of n structure for which component failures are repaired as they occur by a single repair person. System failure occurs when the number of functioning components is reduced by $k + 1$ to $n - k - 1$. The problem may again be represented as a Markov process. In fact, since only adjacent states are accessible in single transitions, it is a birth–death process [60]. This makes analysis possible. However, it is exceedingly complicated to construct the time-dependent transition functions. Birolini [59] gives the expected first passage time to the system down state and the limiting availability. To state his solution, let E_i represent the state that i components are failed, and let δ_i denote the failure intensity for state E_i. Then,

$$\delta_i = k\lambda + (n - k - 1)\lambda_r$$

and the performance measures depend upon the quantity $n - k$. When $n - k = 1$, the expected first passage time to system failure is

$$E[T] = \frac{\delta_0 + \delta_1 + \mu}{\delta_0 \delta_1}$$

and

$$A_\infty = \frac{\mu(\delta_0 + \mu)}{\delta_0 \delta_1 + \mu\delta_0 + \mu^2}$$

and by similar analysis, when $n - k = 2$, the expected first passage time to system failure is

$$E[T] = \frac{\delta_2(\delta_0 + \delta_1 + \mu) + \mu(\delta_0 + \mu) + \delta_0\delta_1}{\delta_0\delta_1\delta_2}$$

and

$$A_\infty = \frac{\mu(\delta_0\delta_1 + \delta_0\mu + \mu^2)}{\delta_0\delta_1\delta_2 + \mu\delta_0\delta_1 + \delta_0\mu^2 + \mu^3}$$

Finally, Birolini provides the general solution for the limiting availability for any value of $n - k$:

$$A_\infty = \frac{1 + \sum_{i=1}^{n-k} \left(\prod_{j=0}^{i-1} \delta_j \right) \Big/ \mu^i}{1 + \sum_{i=1}^{n-k+1} \left(\prod_{j=0}^{i-1} \delta_j \right) \Big/ \mu^i} \tag{11.41}$$

Clearly, more complicated system behavior or even less regular life and repair time distributions may make the analysis of system behavior quite intricate. For the more complicated systems, the general results defined in Section 11.3 provide one approach, and a simulation of the general Markov process is often a worthwhile approach.

11.6 EXERCISES

1. Assume a device has an exponential life distribution with parameter $\lambda_f = 0.005$ and an exponential repair time distribution with parameter $\lambda_r = 0.08$. Plot the point availability function for the device, and plot also the Laplace transform of the availability function, $A^*(s)$.
2. For the device of the previous problem, compute the limiting availability and also the average availability over the interval (50, 150) hours.
3. Reconstruct the numerical results shown in Table 11.1 for the Weibull life distribution and exponential repair time distribution for 50 hours and for 250 hours.
4. Assume the bridge structure is comprised of five identical copies of the device described in Problem 1 above. Plot the point availability function for the system.
5. Suppose a series system is comprised of four components, each of which has an exponential life distribution and an exponential repair time distribution. Assume the parameters for these distributions are $\lambda_{1f} = 0.001$, $\lambda_{1r} = 0.02$, $\lambda_{2f} = 0.003$, $\lambda_{2r} = 0.07$, $\lambda_{3f} = 0.004$, $\lambda_{3r} = 0.03$, $\lambda_{4f} = 0.002$, and $\lambda_{4r} = 0.10$. Plot the point availability function for the system, and compute the limiting system availability.
6. Suppose a device has a Weibull life distribution with parameters $\beta = 2.0$ and $\theta = 200$ hours and is subject to imperfect repair with $p(t) = p = 0.25$. The duration of perfect repair has an exponential distribution with parameter $\lambda_p = 0.08$, and the duration of minimal repair has an exponential distribution with parameter $\lambda_r = 0.02$. Compute the limiting availability for the device.

12

Preventive Maintenance

Manufacturing equipment, electrical power generation stations, airplanes, and automobiles are examples of types of equipment for which operation to failure is usually inefficient and dangerous. While it is true that component and system failures do occur for these equipment types, it is also true that failure events can be very important, and the frequency of failures and also the severity of failures can often be reduced by preventive maintenance.

Preventive maintenance (PM) is the practice of removing a functioning device from operation in order to repair, replace, adjust, test, or simply inspect it. Naturally, the specific preventive maintenance action depends upon the particular system, but all preventive maintenance activity is undertaken with the intent to provide enhanced assurance of function. In some cases, PM involves the replacement of a functioning device prior to failure. Generally, the motivation for replacing a functioning device is that the cost of doing so is small in comparison to the expense of responding to a failure that occurs during device operation — a field failure. Actual failure may imply damage to the device or to other components or equipment. For other cases such as alarm systems, simple inspection and testing are the logical forms of PM. For still other equipment types, adjustment is the appropriate action, while for others lubrication, annealing, or reinforcement is the required forms of service.

In general, performing PM implies some sort of resource commitment, and it is expected that the PM effort will reduce the probability of device failure. The expense of the resources devoted to preventive maintenance is expected to be balanced by the savings in (potentially substantial) expenses associated with in-service failure.

Because of the wide variety of possible applications and the diversity of conceivable policies, numerous models of preventive maintenance have been developed. We will examine several classes of PM strategies and their associated models here. Some of the models are defined in terms of availability. However, because the availability function is usually very difficult to construct and analyze, most of the models are constructed using cost functions for the measure of performance instead of availability. We should recognize that, in general, PM reduces availability even though it also usually reduces cost.

12.1 REPLACEMENT POLICIES

12.1.1 The Elementary Models

Historically, two classes of replacement-type preventive maintenance policies have been defined. They are referred to as "age replacement" and "block replacement" policies. An age replacement policy operates as follows:

> The device is replaced by a new copy upon failure or if it achieves an age equal to the "policy age," τ_a.

In both cases, the device is assumed to be as good as new following replacement so renewal theory-based models may be used to represent device performance. Also, the traditional models were constructed under the assumption that the duration of the replacement activity has the same distribution for both PM and replacement following failure. Replacement or repair following failure is referred to here as "corrective" maintenance, in contrast to PM. The assumption of a common replacement time distribution is used here initially, and a model based on distinct distributions is presented subsequently.

In contrast to an age replacement policy, a block replacement policy is based on scheduled actions rather than on device age. The policy operates as follows:

> The device is replaced by a new copy upon failure or at uniformly spaced time intervals, τ_b, $2\tau_b$, $3\tau_b$, …, where τ_b is called the "policy time."

A slight modification to the policy has been studied by several people. It is that, when the device fails at a time close to the policy time, either the device is left idle until the replacement time, or the replacement time is advanced slightly. As they constitute minor revisions to the models, these adjustments to the policy are not pursued here.

For both age replacement and block replacement, optimal values for the policy times may be determined by analyzing appropriate cost models. This is shown below. First, relationships between the two policies and general patterns of behavior are examined. Start by defining a convenient notation and an ordering relationship among distribution functions. Specifically, let:

$N(t, \tau_a)$ = the number of device failures during an interval $(0, t)$ when the device is operated under an age replacement policy with policy age τ_a.

$N(t, \tau_b)$ = the number of device failures during an interval $(0, t)$ when the device is operated under a block replacement policy with policy time τ_b.

N(t) = the number of device failures during an interval $(0, t)$ when the device is operated to failure with no PM.

$\tilde{N}(t, \tau_a)$ = the number of device replacements (failures and PM) during an interval $(0, t)$ when the device is operated under an age replacement policy with policy age τ_a.

$\tilde{N}(t, \tau_b)$ = the number of device replacements (failures and PM) during an interval $(0, t)$ when the device is operated under a block replacement policy with policy time τ_b.

The general points of interest revolve around the relationships among these quantities. As all are random variables, we use stochastic ordering for which we have:

Defn. 12.1: A random variable X is stochastically greater than or equal to the random variable Y, written $X \overset{st}{\geq} Y$, if

$$\Pr[X > z] \geq \Pr[Y > z] \quad \forall z \tag{12.1}$$

This definition is widely used, but is somewhat counter-intuitive, because it implies that $X \overset{st}{\geq} Y$ when $F_X(z) \leq F_Y(z)$ for all z. The appropriate interpretation is that the distribution on X has its mass concentrated on greater values of the random variable than does the distribution on Y.

Now, using the above definitions, we find that, for all devices with NBU life distributions over all values of t and all choices of τ_a,

$$N(t) \overset{st}{\geq} N_a(t, \tau_a) \tag{12.2}$$

Thus, using an age replacement policy increases the probability that the number of failures during any interval will be small. This provides a reasonable starting justification for the use of PM. Some additional relationships are:

a. For all $t \geq 0$ and $\tau_b \geq 0$, $N(t) \overset{st}{\geq} N_b(t, \tau_b)$ if and only if $F_T(t)$ is NBU.
b. For all $t \geq 0$ and $\tau_a \geq 0$, $N_a(t, \tau_a)$ is stochastically increasing in τ_a if and only if $F_T(t)$ is NBU.
c. For all $t \geq 0$ and $\tau_b \geq 0$, $N_b(t, \tau_b)$ is stochastically increasing in τ_b if and only if $F_T(t)$ is NBU.

With regard to the comparisons of the policies, the relationships are quite interesting. Specifically, for all $t \geq 0$ and for any value assigned to the policy parameters, τ,

$$\begin{aligned} \tilde{N}_a(t, \tau) &\overset{st}{\leq} \tilde{N}_b(t, \tau) \\ N_a(t, \tau) &\overset{st}{\geq} N_b(t, \tau) \end{aligned} \tag{12.3}$$

where the first of these relations holds in general, and the second applies when the life distribution is IFR. Basically, these two inequalities state that block replacement tends to yield more device removals, while age replacement tends to yield more device failures. This seems to conform to intuition, as one has the sense that the block replacement policy involves removal of relatively young copies of the device.

Now, the cost models for these two basic PM strategies are usually formulated without considering the durations of the maintenance tasks. It is considered that the costs adequately capture the implications of failure and of planned replacement, so the total cost per unit time is an informative measure of device performance. It is also considered that the use of PM will reduce the frequency of "field failures" (failures while in operation), and presumably this implies a cost savings. Start with the block replacement policy.

Suppose it is possible to identify the costs of a planned replacement and of field failures, and that these quantities are represented by c_1 and c_2, respectively. Then, a model for the total cost per unit time associated with a block replacement PM strategy is:

$$E[Cost \mid \tau_b] = \frac{c_1 + c_2 M_{F_T}(\tau_b)}{\tau_b} \tag{12.4}$$

The interpretation of this model is that there will be one planned replacement per period at a cost of c_1, and the expected number of failures with corrective replacements per period is given by the renewal function. Each corrective replacement has a cost of c_2. Dividing by the length of the period gives the expected cost per unit time.

Once the cost model is defined, we can use it to determine an optimal choice to the policy time, τ_b, by conventional optimization methods. Taking the derivative,

$$\frac{d}{d\tau_b} E[Cost \mid \tau_b] = \frac{c_2 \tau_b \dfrac{d}{d\tau_b} M_{F_T}(\tau_b) - c_1 - c_2 M_{F_T}(\tau_b)}{\tau_b^2}$$

and equating it to zero, we find that the optimal choice of the policy time is the value for which

$$\tau_b \frac{d}{d\tau_b} M_{F_T}(\tau_b) - M_{F_T}(\tau_b) = {c_1}/{c_2} \tag{12.5}$$

Intuitively, it is appealing that the policy time should depend upon the ratio of preventive to corrective replacement costs. Before analyzing the derivative equation further, observe that the second derivative condition becomes

$$\frac{d^2}{d\tau_b^2} E[Cost \mid \tau_b] = \frac{\begin{pmatrix} c_2\tau_b^3 \dfrac{d^2}{d\tau_b^2} M_{F_T}(\tau_b) - 2c_2\tau_b^2 \dfrac{d}{d\tau_b} M_{F_T}(\tau_b) \\ -2c_1\tau_b - 2c_2\tau_b M_{F_T}(\tau_b) \end{pmatrix}}{\tau_b^4}$$

$$= \frac{c_2}{\tau_b} \frac{d^2}{d\tau_b^2} M_{F_T}(\tau_b)$$

Since c_2 and τ_b are positive, the sign of the second derivative is determined by the life distribution. Recall that a distribution that is NBU has

$$M_{F_T}(x+y) \geq M_{F_T}(x) + M_{F_T}(y)$$

so the slope of the renewal density is positive. Thus, if the device life distribution is NBU, the value of τ_b computed using Expression 12.5 will correspond to a minimum of the expected cost function. Logically, if replacement improves reliability, it is worthwhile, and if replacement does not improve reliability, PM is not appropriate.

Consider an example. Suppose a device has a Weibull life distribution with parameters $\beta = 2.00$ and $\theta = 2000$ hours. Suppose further that the ratio of preventive replacement cost to corrective replacement cost is ρ. Recall that the derivative of the renewal function is the renewal density, so we wish to determine the value of τ_b for which

Table 12.1 Optimal Block Replacement Intervals as a Function of Cost Ratio

ρ	τ_b
0.05	350
0.20	730
0.40	1102
0.50	1282
0.75	1808

$$\tau_b m_{F_T}(\tau_b) - M_{F_T}(\tau_b) = {}^{c_1}\!/_{c_2} = \rho$$

Numerical solution of this expression using Lomnicki's [49] coefficients yields the solutions shown in Table 12.1 as a function of ρ. For the age replacement policy, there is only one replacement per cycle, but it may be either corrective or preventive. If failure occurs before the device has an age of τ_a, replacement will be corrective. If the device survives to age τ_a, replacement is preventive. Thus, the cost per cycle is:

$$c_1\bar{F}_T(\tau_a) + c_2 F_T(\tau_a)$$

and we must distribute this cost over the expected cycle length, which is

$$\tau_a \bar{F}_T(\tau_a) + \int_0^{\tau_a} t f_T(t)dt = \int_0^{\tau_a} \bar{F}_T(t)dt$$

Therefore, the expected cost per unit time is

$$E[Cost \mid \tau_a] = \frac{c_1\bar{F}_T(\tau_a) + c_2 F_T(\tau_a)}{\int_0^{\tau_a} \bar{F}_T(t)dt} \qquad (12.6)$$

Here again, we take the derivative and obtain

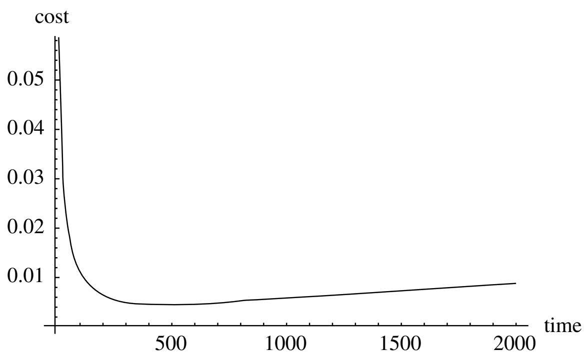


Figure 12.1 Expected cost function for the age replacement.

$$\frac{d}{d\tau_a}E[Cost\,|\,\tau_a]=\frac{\left(\begin{array}{c}\left(-c_1 f_T(\tau_a)+c_2 f_T(\tau_a)\right)\int_0^{\tau_a}\bar{F}_T(t)dt\\ -\left(c_1\bar{F}_T(\tau_a)+c_2 F_T(\tau_a)\right)\bar{F}_T(\tau_a)\end{array}\right)}{\left(\int_0^{\tau_a}\bar{F}_T(t)dt\right)^2}$$

and equating the derivative to zero, we obtain the optimality condition:

$$\bar{F}_T(\tau_a)+z_T(\tau_a)\int_0^{\tau_a}\bar{F}_T(t)dt=\frac{c_2}{c_2-c_1} \tag{12.7}$$

Once again, the solution has the intuitive appeal that the length of the replacement interval depends directly upon a ratio of the replacement costs. Consider the same example as was used for block replacement. A plot of the cost function is shown in Figure 12.1. Solution of Expression 12.7 indicates that the optimal replacement age is τ_a = 460.86 hours.

The block replacement and age replacement policies and models provide an informative starting point for our examination of preventive maintenance. Once the initial models were defined, numerous extensions and improvements were defined, and then more intricate and more efficient policies were defined. We will examine a variety of these extensions in the pages that follow.

12.1.2 Availability Model for Age Replacement

One of the first and most realistic extensions to the basic models is to distinguish between the times required for preventive and corrective maintenance actions. Consideration of the times to perform service implies the analysis of availability. Arguably, the magnitudes of the two cost coefficients in the previous models represent the difference in the service efforts. On the other hand, the availability measures are very appealing descriptors of device performance.

Barlow and Hunter [50] first studied the age replacement policy with distinct service times but took the times to be fixed, so that they were essentially the same as cost factors. Using the fixed times, they obtained the limiting availability but not the point availability function, and they showed that preventive maintenance is not efficient if the cost of PM is the same as the cost of corrective replacement. More recently, Murdock and Nachlas [61] constructed a general availability model for age replacement when the service time distributions are distinct. Examine their model.

To start, let $G_p(t)$ be the distribution function on the times to perform PM, and let $G_c(t)$ be the distribution on corrective maintenance times. Assume that all service implies device renewal, and that the life distribution is represented by $F_T(t)$. Now, each time the renewed device starts operation, two types of operating and service intervals are possible. Operation may end in failure with corrective repair, or it may end after τ_a hours of operation with a preventive maintenance replacement. Thus, the renewal interval may look like either of the ones shown in Figure 12.2. The interval that involves repair following failure occurs with probability $F_T(\tau_a)$, and the one that involves PM occurs with probability $\bar{F}_T(\tau_a)$, so the overall renewal process is a mixture of the two processes created by the operation/repair renewals and the operation/PM renewals. Now, the length of the renewal intervals for the case of operation and repair has distribution $H_{a,c}(t)$, where the distribution is

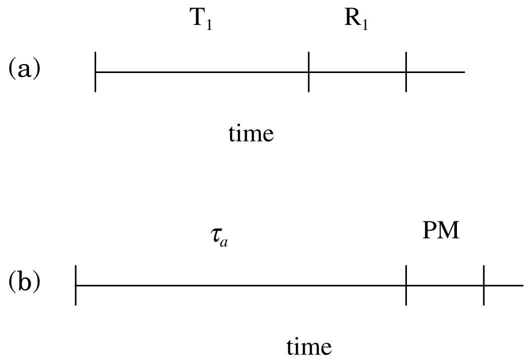


Figure 12.2 (a) Operation to failure with repair. (b) Operation to PM replacement.

the convolution of the repair time distribution with the truncated life distribution:

$$H_{a,c}(t) = \int_0^t G_c(t-u)\frac{f_T(u)}{F_T(\tau_a)}du \tag{12.8}$$

The life distribution is truncated because device operation is not allowed to continue beyond the age replacement policy age. In order to construct this convolution, we will use the partial Laplace transform:

$$f^*_{T,c}(s, \tau_a) = \int_0^{\tau_a} e^{-st}\frac{f_T(t)}{F_T(\tau_a)}dt \tag{12.9}$$

Then, the transform for the density on the length of the interval is

$$h^*_{a,c}(s) = \frac{1}{F_T(\tau_a)} f^*_{T,c}(s, \tau_a) g^*_c(s) \tag{12.10}$$

The renewal function for the process associated with operation and repair intervals is denoted by $M_{H_{a,c}}(t)$.

In the case of the renewal intervals that involve PM, the distribution on the duration of the intervals is

$$H_{a,\,p}(t) = G_p(t - \tau_a) \tag{12.11}$$

as the interval includes operation for τ_a hours followed by a PM activity of random duration. The Laplace transform for the density on the length of this type of interval includes the Heavyside function because of the delayed start of the probability mass. It is

$$h^*_{a,\,p}(s) = e^{-s\tau_a} g^*_p(s) \tag{12.12}$$

The renewal function associated with this process is denoted by $M_{H_{a,\,p}}(t)$, and the density on the duration of the renewal intervals is

$$\begin{aligned} h^*_a(s) &= \bar{F}_T(\tau_a) h^*_{a,\,p}(s) + F_T(\tau_a) h^*_{a,\,c}(s) \\ &= \bar{F}_T(\tau_a) e^{-s\tau_a} g^*_p(s) + f^*_{T,\,c}(s,\,\tau_a) g^*_c(s) \end{aligned} \tag{12.13}$$

The associated renewal function is denoted by $M_{H_a}(t)$.

Given the definition of the renewal function, the availability function is defined in a similar manner to that used to construct Expression 11.13 of Chapter 11.

$$A(t) = \begin{cases} \bar{F}_T(t) + \displaystyle\int_0^t \bar{F}_T(t-u) m_{H_a}(u)du & 0 < t \le \tau_a \\ \displaystyle\int_{t-\tau_a}^t \bar{F}_T(t-u) m_{H_a}(u)du & t > \tau_a \end{cases} \tag{12.14}$$

As with the availability functions of Chapter 11, there are two reasonable approaches to constructing the actual point availability function for a particular choice of life and service time distributions. Direct numerical integration using appropriate numerical approximations can be very effective. Murdock and Nachlas [61] constructed the Laplace transform for Expression 12.14 as

$$A^*(s) = \frac{\bar{F}^*_{T,\,a}(s,\tau_a)}{1 - h^*_a(s)} \tag{12.15}$$

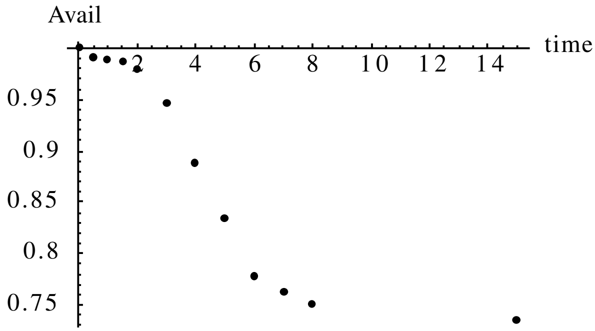


Figure 12.3 Availability function for an example age replacement policy.

They then used numerical inversion algorithms to obtain values for the time-dependent function for several combinations of life and service time distributions. An example of their results for the Weibull life distribution and two different service time distributions is shown in Figure 12.3.

12.1.3 Availability Model for Block Replacement

A similar analysis may be performed for the block replacement policy. Suppose, for example, that the block replacement schedule calls for a device replacement at the end of each month. Then, each month begins with a PM interval of random duration, and this is followed by a period of operation that may in turn be followed by one or more periods of repair and operation. If we formulate our model of this process as the sequence of intervals (PM + T_1), (R_1 + T_2), …, then the failure times form a delayed renewal process. The availability function is thus

$$A(t)=\int_0^t g_p(u)\bar{F}_T(t-u)du+\int_0^t \bar{F}_T(t-u)m_{H_b}(u)du \quad (12.16)$$

In this expression, the renewal density for the delayed renewal process is constructed by first forming the convolutions,

$$h_{b,\,c}(t) = \int_0^t g_c(t-u)\frac{f_T(u)}{F_T(\tau_b - u)}du \tag{12.17}$$

$$h_{b,\,p}(t) = \int_0^t g_p(t-u)\frac{f_T(u)}{F_T(\tau_b - u)}du \tag{12.18}$$

where the terms in the denominators reflect the truncation of the life distribution at the end of the block replacement interval less the time already consumed by replacement. Then, the renewal function is the convolution of $h_{b,p}(t)$ and the renewal density for the ordinary renewal process based on $h_{b,\,c}(t)$. That is,

$$m_{H_b}(t) = h_{b,\,p}(t) * m_{H_{b,\,c}}(t) \tag{12.19}$$

A point here related to the evaluation of this model is that the Laplace transform for the availability function, Expression 12.16, is likely to be completely unmanageable, so direct numerical integration appears to be the most efficient approach to the analysis. The numerical analysis is not too difficult. In addition, for most plausible cases, the first of the two terms of the availability function dominates the other in magnitude. Consider an example. Suppose we have a device with a Weibull life distribution with parameters $\beta = 2.0$ and $\theta = 2000$ and exponential corrective and preventive replacement times with $\lambda_c = 0.005$ and $\lambda_p = 0.025$. If we take $\tau_b = 730$ hours, the availability function will be as is shown in Figure 12.4. Finally, note that Expression 12.16 applies for t in the interval $[0, \tau_b]$, or to say it another way, it is defined relative to the length of the block replacement policy time.

12.1.4 Availability Model for Opportunistic Age Replacement

There are a number of alternate replacement-type preventive maintenance policies that may be defined. One particular possibility is the use of an "opportunistic" preventive maintenance strategy. This is an appealing and quite realistic policy in which the equipment down state is exploited to perform additional maintenance tasks. Clearly such a policy yields

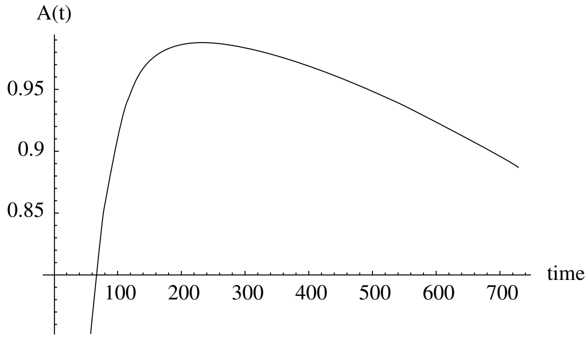


Figure 12.4 Availability function for an example block replacement policy.

efficiencies. It is also commonly used. People often choose to combine preventive maintenance actions in order to reduce the number of times a machine or vehicle must be removed from service. However, models for opportunistic maintenance have not been developed until recently because of the complexity of the probability analysis required. Barlow and Proschan [11] do give the limiting availability for the opportunistic policy, but it was only recently that Degbotse and Nachlas [62] constructed the time-dependent availability function for a general model of opportunistic age replacement. Examine that model.

To start, assume, as before, that replacement implies renewal. Then, consider that any system may be viewed as comprised of two components — one of interest and the balance of the system. The policy operates as follows:

a. If either component fails, it is replaced and the other component does not age during that replacement.
b. If either component attains an age corresponding to its age replacement policy age, it is replaced preventively and the other component does not age during that activity.
c. If at any time the system is "down" in order to replace one of the components, the other component will also

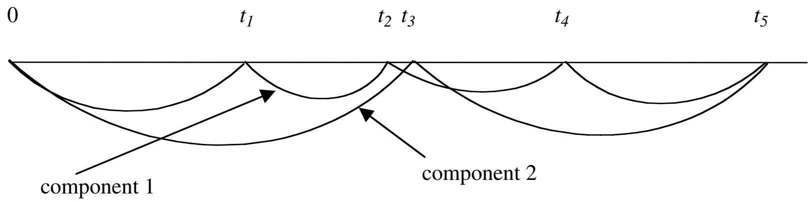


Figure 12.5 Illustration of a nested renewal process for opportunistic replacement. (From Degbotse, A.T. and J.A. Nachlas, "Use of Nested Renewals to Model Availability Under Opportunistic Maintenance Policies," *Proc. of the Annual Reliability and Maintainability Symposium*, pp. 344–350, IEEE, 2003. With permission.)

be replaced if its age exceeds the opportunistic replacement policy age ω.

A model for this policy is rather intricate and requires considerable notation. Let τ_{ai} represent the age replacement policy age for component i and ω_{ai} the opportunistic replacement policy age for component i. Necessarily, $\omega_{ai} < \tau_{ai}$. Also, let $F_{T_i}(t)$ represent the life distribution function for component i and $G_{T_i}(t)$ represent the distribution on the time to replace component i. These quantities are used to define a nested renewal process that represents the operation of the components and, thus, of the system.

An illustration of the nested renewal process is shown in Figure 12.5. Notice that the renewal points for "component 1," points t_1, t_2, t_4, and t_5 are "nested" within those for "component 2," points t_3 and t_5. To describe operating profiles that correspond to nested renewal processes, we define two classes of operating intervals, major intervals and minor intervals, and we further distinguish between initial and general minor intervals. A "major interval" is a period of system operation and repair that starts with both components being new and ends with the completion of a service period in which both components are replaced (renewed). Thus, a major interval is a system renewal period, and a sequence of major intervals forms a renewal process. In Figure 12.5, the interval $(0, t_5)$ is a major interval.

In contrast, a "minor interval" is a period between two successive restarts of the system following replacement of a specific one of the components. Since one may use either of the two components to delimit a minor interval, we arbitrarily use "component 2." Thus, minor intervals are the periods between component 2 restarts. The sequence of minor intervals thus forms a renewal process for component 2 but not for the system.

Next, distinguish further between an "initial minor interval," which starts with both components being new, and a "general minor interval," which starts with component 2 being new and component 1 being used. In Figure 12.5, the interval $(0, t_3)$ is an initial minor interval, as both components are new when it starts. It ends with the system restart following replacement of component 2. During the interval, component 1 has been replaced several times. The interval (t_3, t_5) is a general minor interval, because it starts with component 1 being used and component 2 being new. A major interval will be comprised of an initial minor interval and some number (possibly zero) of general minor intervals.

Now, given the life and replacement time distributions, we define several quantities that help describe system operation. We focus on cases in which system renewal does not occur, and these lead to expressions that represent system renewal and availability. Let:

$H_I(t, k)$ = the cumulative distribution function on the accumulated operating time during an initial minor interval that ends with the replacement of component 2 following k component 1 replacements and without system renewal.

$H_G(t, k)$ = the cumulative distribution function on the accumulated operating time during a general minor interval that ends with the replacement of component 2 following k component 1 replacements and without system renewal.

Denote the density functions as $h_I(t, k)$ and $h_G(t, k)$. The functions H_I and H_G are the key building blocks of the model of system behavior. They are constructed using the life and the residual life distributions. That construction is postponed for now so that the general model structure can be defined without treating special cases.

Since a minor interval that ends without system renewal includes k component 1 replacements and one component 2 replacement, the density function on the total repair time during a minor interval is

$$g_T(t, k) = \int_0^t g_1^{(k)}(x) g_2(t - x) dx \tag{12.20}$$

where $g_1^{(k)}(t)$ is the k-fold convolution of the replacement time density.

The density function on the length of a minor interval is the convolution of the operating and the repair time densities. That is, for

$Q_I(t)$ = the distribution function on the length of an initial minor interval that ends without system renewal.
$Q_G(t)$ = the distribution function on the length of a general minor interval that ends without system renewal.

we obtain

$$q_I(t) = \sum_{k=0}^{\infty} \int_0^t h_I(x, k) g_T(t - x, k) dx \tag{12.21}$$

$$q_G(t) = \sum_{k=0}^{\infty} \int_0^t h_G(x, k) g_T(t - x, k) dx \tag{12.22}$$

$$Q_I(t) = \int_0^t q_I(x) dx \tag{12.23}$$

$$Q_G(t) = \int_0^t q_G(x) dx \tag{12.24}$$

Corresponding relationships are defined for the minor intervals that end with system renewal. Let

$U_I(t)$ = the distribution function on the length of an initial minor interval that ends with system renewal.

$U_G(t)$ = the distribution function on the length of a general minor interval that ends with system renewal.

To construct these functions, we use

$V_{I1}(t, k)$ = the cumulative distribution function on the accumulated operating time during an initial minor interval that ends with system renewal due to the replacement of both components following a component 1 failure.

$V_{I2}(t, k)$ = the cumulative distribution function on the accumulated operating time during an initial minor interval that ends with system renewal due to the replacement of both components following a component 2 failure.

$V_{G1}(t, k)$ = the cumulative distribution function on the accumulated operating time during a general minor interval that ends with system renewal due to the replacement of both components following a component 1 failure.

$V_{G2}(t, k)$ = the cumulative distribution function on the accumulated operating time during a general minor interval that ends with system renewal due to the replacement of both components following a component 1 failure.

Finally, define

$Z_I(t)$ = the distribution function on the length of an initial minor interval that ends with or without system renewal.

$Z_G(t)$ = the distribution function on the length of a general minor interval that ends with or without system renewal.

Naturally,

$$Z_I(t) = \sum_{k=0}^{\infty} H_I(t, k) + \sum_{k=0}^{\infty} V_{I1}(t, k) + \sum_{k=0}^{\infty} V_{I2}(t, k) \qquad (12.25)$$

and

$$Z_G(t) = \sum_{k=0}^{\infty} H_G(t,k) + \sum_{k=0}^{\infty} V_{G1}(t,k) + \sum_{k=0}^{\infty} V_{G2}(t,k) \quad (12.26)$$

The functions defined above can be used to construct two availability measures. First, we obtain the availability within a sequence of minor intervals that do not yield system renewal, and we then incorporate that availability measure in an overall system availability function. The renewal density for the minor intervals without having system renewal is given by

$$m_Q(t) = \sum_{n=0}^{\infty} \int_0^t q_I(x) q_G^{(n)}(t-x)dx \quad (12.27)$$

This represents the probability that a minor interval starts at any time. For the sequence of minor intervals, the system availability is the probability that the system is operating at any point in time and is thus given by

$$A(t) = \bar{Z}_I(t) + \int_0^t \bar{Z}_G(t-x) m_Q(x)dx \quad (12.28)$$

where $\bar{Z}_I(t) = 1 - Z_I(t)$ represents the probability that the length of the operating period in an initial minor interval exceeds t.

The system-level availability starts with the distribution function for the lengths of the major intervals, $\Phi(t)$, and the corresponding renewal density, $m_\Phi(t)$. The distribution on the lengths of the major intervals is

$$\Phi(t) = U_I(t) + \int_0^t m_Q(x) U_G(t-x)dx \quad (12.29)$$

as either an initial or a general minor interval may be the last minor interval in a major interval. Given this definition, the system level availability function is

$$A_S(t) = A(t) + \int_0^t A(x) m_\Phi(t-x)dx \tag{12.30}$$

This form of the system-level availability function reflects the nesting of the minor intervals within the major intervals.

Next, observe that we can apply the general model structure to four problem classes, and that the stepwise application over the four classes will promote understanding of the model structure and will allow us to connect our availability results to the few others that are available.

Recall that the two components each have an age replacement policy age τ_{ai} and an opportunistic replacement age ω_{ai}. The four cases are defined by the assumed magnitudes of those policy parameters. Specifically, we consider (1) $\tau_{a1} = \tau_{a2} = \omega_{a1} = \omega_{a2} = \infty$, (2) $\tau_{a1} = \tau_{a2} = \infty$ with ω_{a1} and ω_{a2} finite, (3) $\tau_{a1} = \infty$ with τ_{a2}, ω_{a1} and ω_{a2} finite, and (4) all policy ages finite. For each case, the specification of the policy ages leads to the definition of the functions H_I and H_G. The first case is a pure failure model.

12.1.4.1 A Failure Model

When we take all of the policy ages to be infinite, we say that there will be no PM and no opportunistic replacement. Components will only be replaced upon failure. This implies that system renewal will never occur. Refer again to Figure 12.5 and note that, for an initial minor interval, the probability that component 2 fails on or before t time units of operation, during which component 1 fails k times, is

$$H_I(t, k) = \begin{cases} \bar{F}_1(t) F_2(t) & k = 0 \\ \int_0^t \bar{F}_1(t-u) f_1^{(k)}(u) F_2(t) du & k \geq 1 \end{cases} \tag{12.31}$$

The associated density function is

$$h_I(t, k) = \begin{cases} \bar{F}_1(t) f_2(t) & k = 0 \\ \int_0^t \bar{F}_1(t-u) f_1^{(k)}(u) f_2(t) du & k \geq 1 \end{cases} \tag{12.32}$$

The corresponding forms for the general minor interval are

$$H_G(t,k)=\begin{cases}\bar{\tilde{F}}_1(t)F_2(t) & k=0\\ \int_0^t \bar{F}_1(t-u)\tilde{f}_1^{(k)}(u)F_2(t)du & k\geq 1\end{cases} \tag{12.33}$$

and

$$h_G(t,k)=\begin{cases}\bar{\tilde{F}}_1(t)f_2(t) & k=0\\ \int_0^t \bar{F}_1(t-u)\tilde{f}_1^{(k)}(u)f_2(t)du & k\geq 1\end{cases} \tag{12.34}$$

where it should be particularly noted that $\bar{\tilde{F}}_1(t)$ denotes the survivor function for the residual life distribution, $\tilde{F}_1(t)$. The residual life distribution is used because the general minor intervals begin with component 1 being used. Note further that the convolution, $\tilde{f}_1^{(k)}(t)$, includes a first operating period during which component 1 is used followed by $k - 1$ periods in which component 1 is new at the start.

In general, it is difficult to specify the residual life distribution for a process such as the one studied here. We use the approximation to the residual life distribution defined by Cox [44]:

$$\tilde{F}_T(t)=\frac{1}{\mu_T}\int_0^t \bar{F}_T(x)dx \tag{12.35}$$

where μ_T is the mean of the original life distribution.

Next, we construct

$$\begin{aligned} q_I(t)=&\int_0^t f_2(u)\bar{F}_1(u)g_2(t-u)du\\ &+\sum_{k=1}^{\infty}\int_0^t\left(f_2(x)\int_0^x \bar{F}_1(x-u)f_1^{(k)}(u)du\right)\\ &\left(\int_0^{t-x} g_1^{(k)}(w)g_2(t-x-w)dw\right)dx \end{aligned} \tag{12.36}$$

and

$$q_G(t) = \int_0^t f_2(u)\bar{\bar{F}}_1(u) g_2(t-u)du + \sum_{k=1}^{\infty} \int_0^t \left(f_2(x) \int_0^x \bar{F}_1(x-u) \tilde{f}_1^{(k)}(u)du \right) \left(\int_0^{t-x} g_1^{(k)}(w) g_2(t-x-w)dw \right) dx \tag{12.37}$$

These forms are substituted into Expression 12.27 to obtain the renewal density for minor intervals. The distributions on total operating time during minor intervals are

$$Z_I(t) = \sum_{k=0}^{\infty} H_I(t, k) = \sum_{k=0}^{\infty} \int_0^t f_2(x) \int_0^x \bar{F}_1(x-u) f_1^{(k)}(u)dudx = F_2(t) \tag{12.38}$$

and

$$Z_G(t) = \sum_{k=0}^{\infty} H_G(t, k) = \sum_{k=0}^{\infty} \int_0^t f_2(x) \int_0^x \bar{F}_1(x-u) \tilde{f}_1^{(k)}(u)dudx = F_2(t) \tag{12.39}$$

which is an appealing and rather intuitive result. An important point here is that the simplification of the general model to the simple case of the failure process provides some validation of the general formulation. The expressions for $u_I(t)$ and $u_G(t)$ are constructed in the same fashion, as is shown shortly.

Returning to the functions $Q_I(t)$ and $Q_G(t)$, note that $Q_I(\infty)$ and $Q_G(\infty)$ are the probabilities for the two types of minor intervals that opportunistic replacement does not occur. If

$Q_I(\infty) < 1$ and $Q_G(\infty) < 1$, then there is a positive probability that an opportunistic replacement and, hence, a system renewal occurs. In this case, the minor intervals are transient. On the other hand, if $Q_I(\infty) = 1$ and $Q_G(\infty) = 1$, then there is no system renewal, and the minor intervals are recurrent. For the failure model, $Q_I(\infty) = 1$ and $Q_G(\infty) = 1$ and also $Z_I(\infty) = 1$ and $Z_G(\infty) = 1$. Hence, the minor intervals are recurrent. As a result, the system availability function reduces to the within minor interval availability defined in Expression 12.28. Further, based on Expressions 12.38 and 12.39, the system availability is

$$A(t) = \bar{F}_2(t) + \int_0^t \bar{F}_2(t-x)m_Q(x)dx \tag{12.40}$$

As usual, depending upon the life and repair time distributions, evaluation of the availability function can be rather intricate. As noted previously, Barlow and Proschan [11] present the limiting availability for the failure model as

$$A_\infty = \left(1 + \frac{\nu_1}{\mu_1} + \frac{\nu_2}{\mu_2}\right)^{-1} \tag{12.41}$$

in which ν_i is the mean of the repair time distribution $G_i(t)$, and μ_i is the mean of the life distribution for component i. Using the derivatives of the Laplace transforms of Expression 12.40 yields the same result.

12.1.4.2 An Opportunistic Failure Replacement Policy

Next, suppose that each of the components is replaced either upon failure or if it has age ω_{ai} when system operation is interrupted by the failure of the other component. We first construct expressions for the probabilities that various sample paths yield restarts for the minor intervals without system renewal. We must perform the construction relative to the magnitudes of ω_{a1} and ω_{a2}. We could consider both cases, but for explanation, let us simply take $\omega_{a1} < \omega_{a2}$. The corresponding analysis and results for the case in which $\omega_{a1} > \omega_{a2}$ are implied by this analysis. To start, note that, when $k = 0$,

$$h_I(t, 0) = f_2(t)\bar{F}_1(t) \tag{12.42}$$

provided $t \le \omega_{a1}$. If $k \ge 1$, the cases that imply no system renewal are:

a. Component 2 fails at time t where $0 \le t < \omega_{a1}$.
b. Component 2 fails at time t where $\omega_{a1} \le t < \omega_{a2}$, and the k^{th} component 1 failure occurred at time x, where $t - x < \omega_{a1}$. That is, $x > t - \omega_{a1}$.
c. Component 2 fails at time t where $t \ge \omega_{a1}$, and the k^{th} component 1 failure occurred at time x, where $x < \omega_{a2}$ and $t - x < \omega_{a1}$. That is, $t - \omega_{a1} < x < \omega_{a2}$, and by implication, $t \le \omega_{a1} + \omega_{a2}$.

Combining these cases, the realization of Expression 12.33 is

$$h_I(t, k) = \begin{cases} f_{T2}(t)\int_0^t \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)du & 0 \le t < \omega_{a1} \\ f_{T2}(t)\int_{t-\omega_{a1}}^t \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)du & \omega_{a1} \le t < \omega_{a2} \\ f_{T2}(t)\int_{t-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)du & \omega_{a2} \le t < \omega_{a1} + \omega_{a2} \end{cases} \tag{12.43}$$

The expressions in 12.42 and 12.43 are used in 12.21 to obtain the density on the lengths of the initial minor intervals.

Next, we consider the general minor intervals that end without system renewal. The reasoning and, hence, construction for the general minor intervals are identical to that for the initial minor intervals, except that component 1 is used and is subject to a residual life distribution at the start of a general minor interval. To reflect the age of component 1 at the start of a general minor interval, we take the component 1 age to be the average backward recurrence time based on the life distribution, $F_{T1}(t)$. Denote that age as a_1, and note that Cox [44] shows that age to be

$$a_1 = \frac{\mu_{T1}^2 + \sigma_{T1}^2}{2\mu_{T1}} \tag{12.44}$$

Given the average starting age a_1,

$$h_G(t,0) = f_{T2}(t)\bar{\tilde{F}}_{T1}(t)$$

$$0 \le t \;< \omega_{a1} - a_1 \tag{12.45}$$

and

$$h_G(t,k) = \begin{cases} f_{T2}(t)\int_0^t \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & 0 \le t < \omega_{a1} \\ f_{T2}(t)\int_{t-\omega_{a1}}^t \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a1} \le t < \omega_{a2} \\ f_{T2}(t)\int_{t-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a2} \le t < \omega_{a1} + \omega_{a2} \end{cases} \tag{12.46}$$

Again, these expressions are combined with the total repair time to determine the density on the general minor interval length.

Observe that the conditions under which renewal does not occur correspond to a specific set of sample paths. The set of possible sample paths that is the complement of the set described above defines the cases in which system renewal does occur. Considering initial minor intervals, we note that system renewal will occur as a result of a component 2 replacement if $k = 0$ and component 2 fails at time t where $\omega_{a1} \le t$. That is,

$$v_{I2}(t,0) = f_{T2}(t)\bar{F}_{T1}(t)$$

$$\omega_{a1} \le t \tag{12.47}$$

and a component 2 failure will precipitate system renewal for the sample paths having a component 2 failure time of t where

a. $\omega_{a1} \le t < \omega_{a2}$, and the k^{th} component 1 failure time occurred at x, where $0 \le x \le t - \omega_{a1}$

or

b. $\omega_{a2} \le t < \omega_{a1} + \omega_{a2}$, the k^{th} component 1 failure occurred at x, where $0 \le x \le t - \omega_{a1}$, and it is also the case that $x < \omega_{a2}$

or

c. $\omega_{a1} + \omega_{a2} < t$, and the k^{th} component 1 failure occurred at x, where $x < \omega_{a2}$

while it will be the k^{th} component 1 failure that precipitates the system renewal at time t if that failure occurs at time t where $\omega_{a2} \le t < \omega_{a1} + \omega_{a2}$, and the $k - 1^{st}$ failure occurred at time x where $x < \omega_{a2}$. or if that failure occurs at time $t > \omega_{a1} + \omega_{a2}$, and the $k - 1^{st}$ failure occurred at time $x < \omega_{a2}$. These cases exhaust the complement of the set of sample paths that do not yield system renewal at the end of an initial minor interval. Algebraically,

$$v_{I2}(t,k) = \begin{cases} f_{T2}(t)\displaystyle\int_0^{t-\omega_{a1}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)\,du & \omega_{a1} \le t < \omega_{a2} \\ f_{T2}(t)\displaystyle\int_0^{t-\omega_{a1}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)\,du & \omega_{a2} \le t < \omega_{a1} + \omega_{a2} \\ f_{T2}(t)\displaystyle\int_0^{\omega_{a2}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)\,du & \omega_{a1} + \omega_{a2} \le t \end{cases} \quad (12.48)$$

and

$$v_{I1}(t,k) = \bar{F}_{T2}(t)\int_0^{\omega_{a2}} f_{T1}(t-u) f_{T1}^{(k-1)}(u)\,du \qquad \omega_{a2} \le t \quad (12.49)$$

To combine the operating and repair times, we observe that, when a component 2 failure precipitates system renewal, the accumulated repair time is $g_T(t, k+1)$, and in contrast, when a component 1 failure precipitates system renewal, the accumulated repair time is $g_T(t, k)$. Thus, the density function for the duration of an initial minor interval that ends with system renewal is

$$u_I(t) = \sum_{k=0}^{\infty}\int_0^t v_{I1}(x,k) g_T(t-x,k)\,dx + \sum_{k=0}^{\infty}\int_0^t v_{I2}(x,k) g_T(t-x,k+1)\,dx \quad (12.50)$$

For the general minor intervals that end with system renewal, we again consider the sample paths that form the complement of the set of paths that do not yield renewal, so

$$v_{G2}(t,0) = f_{T2}(t)\bar{\tilde{F}}_{T1}(t)$$
$$\omega_{a1} - a_1 \le t \qquad (12.51)$$

and

$$v_{G2}(t,k) = \begin{cases} f_{T2}(t)\int_0^{t-\omega_{a1}} \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a1} \le t < \omega_{a1} + \omega_{a2} \\ f_{T2}(t)\int_0^{\omega_{a2}} \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a1} + \omega_{a2} \le t \end{cases} \qquad (12.52)$$

and

$$v_{I1}(t,k) = \bar{F}_{T2}(t)\int_0^{\omega_{a2}} f_{T1}(t-u)\tilde{f}_{T1}^{(k-1)}(u)du \qquad \omega_{a2} \le t \quad (12.53)$$

Then, the density on the length of the general minor intervals that end with system renewal is

$$u_G(t) = \sum_{k=0}^{\infty}\int_0^t v_{G1}(x,k)g_T(t-x,k)dx + \sum_{k=0}^{\infty}\int_0^t v_{G2}(x,k)g_T(t-x,k+1)dx \qquad (12.54)$$

Having enumerated and represented all of the sample paths for the minor intervals, the system availability is determined by the successive application of Equations 12.25 through 12.30. That is, the specification of any particular realization of the general model structure hinges upon the careful construction of the quantities $q_I(t)$ and $q_G(t)$, using $h_I(t,k)$ and $h_G(t,k)$, and the quantities $u_I(t)$ and $u_G(t)$, using $v_{I1}(t,k)$, $v_{I2}(t,k)$, $v_{G1}(t,k)$, and $v_{G2}(t,k)$. Once these probabilities have been obtained, Equations 12.25 through 12.30 accumulate their content to provide the system availability measure.

To complete this model, we observe that there is a finite probability of system renewal when the system is subjected to an opportunistic failure replacement policy, and this is confirmed for the stated expressions in that $Q_I(\infty) < 1$ and

$Q_G(\infty) < 1$. Also, one can obtain a partial validation of the stated expressions by noting that taking $\omega_{a1} = \omega_{a2} = \infty$ yields the equations for the failure replacement model. Finally, note that it is possible to obtain the limiting system availability by taking derivatives of the pertinent Laplace transforms. Doing this yields

$$A_S(\infty) = \frac{\mu_{ZI}(1 - Q_G(\infty)) + \mu_{ZG} Q_I(\infty)}{(\mu_{UI} + \mu_{QI} + (\nu_1 + \nu_2))(1 - Q_G(\infty)) + (\mu_{UG} + \mu_{QG}) Q_I(\infty)} \quad (12.55)$$

where the mean values are identified by their subscripts, which correspond to the distributions to which they apply. Note that this result conforms to that of Barlow and Proschan [11].

12.1.4.3 A Partial Opportunistic Age Replacement Policy

A partial opportunistic age replacement policy is one in which both ω_{a1} and ω_{a2} are finite, but only one of τ_{a1} and τ_{a2} is. Arbitrarily, let τ_{a2} be finite, and assume $\omega_{a1} < \omega_{a2}$. Then, system renewal occurs when either component 1 fails and component 2 has age of at least ω_{a2}, or component 2 fails or achieves an age of τ_{a2}, and the age of component 1 is at least ω_{a1}. Keep in mind the fact that the analysis is slightly different depending upon whether $\tau_{a2} < \omega_{a1} + \omega_{a2}$ or $\tau_{a2} > \omega_{a1} + \omega_{a2}$. For now, we consider only the first of those cases.

As with the previous model, consider first the initial minor interval that ends without system renewal. Let t denote the time of the component 2 failure (if there is one), and let x denote the time of the k^{th} component 1 failure (if there is one). The sample paths for which there is no system renewal are:

$$S_{k=0} = \left\{(t \mid 0 \le t < \omega_1)\right\}$$

and

$$S_{k\ge 1} = \left\{ t,\, x \,\middle|\, \begin{cases} 0 \le t < \omega_{a1}, 0 \le x < t \\ \omega_{a1} \le t < \omega_{a2}, t - \omega_{a1} \le x < t \\ \omega_{a2} \le t < \tau_{a2}, t - \omega_{a1} \le x < \omega_{a2} \\ t = \tau_{a2}, \tau_{a2} - \omega_{a1} \le x < \omega_{a2} \end{cases} \right\}$$

These sets imply that

$$h_I(t,0) = f_{T2}(t)\bar{F}_{T1}(t)$$

$$0 \le t < \omega_{a1} \tag{12.56}$$

and

$$h_I(t,k) = \begin{cases} f_{T2}(t)\int_0^t \bar{F}_{T1}(t-u)f_{T1}^{(k)}(u)du & 0 \le t < \omega_{a1} \\ f_{T2}(t)\int_{t-\omega_{a1}}^t \bar{F}_{T1}(t-u)f_{T1}^{(k)}(u)du & \omega_{a1} \le t < \omega_{a2} \\ f_{T2}(t)\int_{t-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u)f_{T1}^{(k)}(u)du & \omega_{a2} \le t < \tau_{a2} \\ \bar{F}_{T2}(t)\int_{\tau_{a2}-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u)f_{T1}^{(k)}(u)du & t = \tau_{a2} \end{cases} \tag{12.57}$$

The same sample paths apply to the general minor interval except that the initial age must be considered when $k = 0$. Thus,

$$h_G(t,0) = f_{T2}(t)\bar{F}_{T1}(t)$$

$$0 \le t < \omega_{a1} - a_1 \tag{12.58}$$

and

$$h_G(t,k) = \begin{cases} f_{T2}(t)\int_0^t \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & 0 \le t < \omega_{a1} \\ f_{T2}(t)\int_{t-\omega_{a1}}^t \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a1} \le t < \omega_{a2} \\ f_{T2}(t)\int_{t-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & \omega_{a2} \le t < \tau_{a2} \\ \bar{F}_{T2}(t)\int_{\tau_{a2}-\omega_{a1}}^{\omega_{a2}} \bar{F}_{T1}(t-u)\tilde{f}_{T1}^{(k)}(u)du & t = \tau_{a2} \end{cases} \tag{12.59}$$

In order to identify the probability functions for the system renewal cases, we take the complements of the sets $S_{k=0}$ and $S_{k\ge1}$. Clearly,

$$\bar{S}_{k=0} = \{(t \mid \omega_{a1} \le t)\}$$

and

$$\bar{S}_{k\ge1} = \left\{ t, x \,\middle|\, \begin{Bmatrix} \omega_{a1} \le t < \omega_{a2},\ x \le t - \omega_{a1} \\ \omega_{a2} \le t < \tau_{a2},\ x \le t - \omega_{a1} \\ \omega_{a2} \le t < \tau_{a2},\ x \ge \omega_{a2} \\ t = \tau_{a2},\ x \le \tau_{a2} - \omega_{a1} \end{Bmatrix} \right\}$$

so

$$v_{I2}(t, 0) = f_{T2}(t)\bar{F}_{T1}(t)$$

$$\omega_{a1} \le t \tag{12.60}$$

and

$$v_{I2}(t, k) = \begin{cases} f_{T2}(t)\displaystyle\int_0^{t-\omega_{a1}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)du & \omega_{a1} \le t < \tau_{a2} \\ \bar{F}_{T2}(t)\displaystyle\int_0^{\tau_{a2}-\omega_{a1}} \bar{F}_{T1}(t-u) f_{T1}^{(k)}(u)du & t = \tau_{a2} \end{cases} \tag{12.61}$$

and

$$v_{I1}(t, k) = \bar{F}_{T2}(t)\int_0^{\tau_2} f_{T1}(t-u) f_{T1}^{(k-1)}(u)du \qquad \omega_{a2} \le t < \tau_{a2} \tag{12.62}$$

As with the nonrenewal case, the general minor intervals are similar to the initial minor intervals, so,

$$v_{G2}(t, 0) = f_{T2}(t)\bar{\tilde{F}}_{T1}(t)$$

$$\omega_{a1} - \mathrm{a}_1 \le t \tag{12.63}$$

and

$$v_{G2}(t, k) = \begin{cases} f_{T2}(t)\displaystyle\int_0^{t-\omega_{a1}} \bar{F}_{T1}(t-u) \tilde{f}_{T1}^{(k)}(u)du & \omega_{a1} \le t < \tau_{a2} \\ \bar{F}_{T2}(t)\displaystyle\int_0^{\tau_{a2}-\omega_{a1}} \bar{F}_{T1}(t-u) \tilde{f}_{T1}^{(k)}(u)du & t = \tau_{a2} \end{cases} \tag{12.64}$$

and

$$v_{G1}(t, k) = \bar{F}_{T2}(t)\int_0^{\omega_{a2}} f_{T1}(t-u)\tilde{f}_{T1}^{(k-1)}(u)du \qquad \omega_{a2} \le t < \tau_{a2} \quad (12.65)$$

Once the basic densities on interval duration have been defined relative to the cases that do and do not yield system renewal, the calculation of availability measures follows the previously defined format.

The densities $u_I(t)$ and $u_G(t)$ are computed as given in Expressions 12.50 and 12.54, respectively. These are then used in Equations 12.25 through 12.30 to obtain the system availability measure. Observe that $Q_I(\infty) < 1$ and $Q_G(\infty) < 1$, so system renewals do occur for this model, and Equation 12.55 again gives the limiting system availability. Also, setting $\tau_{a2} = \infty$ causes the model to reduce to the opportunistic failure replacement model as it should.

12.1.4.4 A Full Opportunistic Age Replacement Policy

In a full opportunistic age replacement policy, all of the policy parameters have finite values and both planned replacements and failure events have the potential to precipitate a system renewal. A model of system operation under a full opportunistic age replacement policy is defined in the same manner as the previously constructed models. In fact, the previous models subsume most of the model forms for the full opportunistic replacement case.

Note that each of the models defined previously are constructed in terms of the relative magnitudes of the policy parameters. If we further assume that $\tau_{a1} \ge \tau_{a2}$, then the models for the partial opportunistic age replacement policy apply to the full policy case, because the value τ_{a2} bounds the length of the minor intervals such that age replacement of component 1 cannot occur. Thus, for the full opportunistic age replacement policy, we need consider only cases in which $\tau_{a1} < \tau_{a2}$. However, we note further that the assignment of the indices to the components is arbitrary, so either component may be considered component 1, and it will always be the case that one of the age replacement times limits the duration of the minor intervals. Hence, the models defined above exhaust the "practically interesting" cases and are sufficient to study all of the opportunistic replacement policies enumerated here.

To close the discussion of the construction of the models, observe that the four models are mutually consistent, in that each is successively a generalization of the preceding model.

12.1.4.5 Analysis of the Opportunistic Replacement Models

As we have seen previously, availability models tend to be difficult to analyze, particularly in closed algebraic form. The above models of opportunistic replacement share the attribute of being relatively intractable. However, as with most availability models, we can obtain the Laplace transform of most of the time-dependent measures and can then invert the transform numerically. We can also use the transform to obtain the limiting availability. At the very least, this permits verification of the stability of the system and validation of the models.

It is not necessary to show all of the available results here, but a few key results will help to illustrate the power of the nested renewal concept. To start, we note that

$$q_I^*(s) = \sum_{k=0}^{\infty} h_I^*(s, k) g_T^*(s, k) \tag{12.66}$$

is well defined, so

$$Q_I(\infty) = q_I^*(s)\Big|_{s=0} \tag{12.67}$$

and the same applies to $Q_G(\infty)$. Substituting into Equation 12.27 yields

$$m_Q^*(s) = \frac{q_I^*(s)}{1 - q_G^*(s)} \tag{12.68}$$

Then, using the derivatives of the Laplace transforms for Equation 12.26,

$$\mu_\Phi = -\frac{d}{ds}\Phi^*(s) = -\frac{d}{ds}\left[u_I^*(s) + m_Q^*(s)u_G^*(s)\right]$$
$$= \mu_{UI} + \mu_{QI} + (\mu_{UG} + \mu_{QG})\frac{Q_I(\infty)}{1 - Q_G(\infty)} \tag{12.69}$$

and this leads directly to the limiting availability of Expression 12.55. The intermediate steps are to take

$$A_S^*(s) = A^*(s)\left[1 + m_Q^*(s)\right] \tag{12.70}$$

and

$$A^*(s) = \bar{Z}_I^*(s) + \bar{Z}_G^*(s)m_Q^*(s) \tag{12.71}$$

along with the standard definition of a renewal function for $m_\Phi^*(s)$ to obtain

$$A_S^*(s) = \frac{\bar{Z}_I^*(s)(1 - q_G^*(s)) + \bar{Z}_G^*(s)q_I^*(s)}{(1 - \phi^*(s))(1 - q_G^*(s))} \tag{12.72}$$

Note that all of these results apply regardless of the choices of life and repair time distributions.

Next, to illustrate the construction of numerical results, assume that the distributions on the replacement times are exponential and are

$$g_1(t) = \lambda_1 e^{-\lambda_1 t}$$

and

$$g_2(t) = \lambda_2 e^{-\lambda_2 t}$$

respectively, and that the life distributions are Weibull and are

$$F_{Tj}(t) = 1 - \mathrm{e}^{-\left(t/\theta_j\right)^{\beta_j}}$$

Table 12.2 Representative Limiting Availability Values

γ	$m_1 = 1, m_2 = 1$	$m_1 = 1, m_2 = 2$	$m_1 = 2, m_2 = 1$	$m_1 = 2, m_2 = 2$
0.2	0.47	0.51	0.48	0.54
0.4	0.49	0.52	0.64	0.66
0.6	0.54	0.53	0.62	0.65
0.8	0.63	0.53	0.55	0.61

For the purposes of illustration, and recognizing that scale is arbitrary, take $\beta_1 = \beta_2 = 2.0$ with $\theta_1 = 1.5$, $\theta_2 = 1.0$, $\lambda_2 = 0.4$, and $\lambda_1 = 0.667$. Then, take the age replacement policy ages to be integer multiples of the mean of the life distributions, so that

$$\tau_{aj} = m\mu_j$$

and take the opportunistic age replacement policy ages to be fractions of the age replacement policy ages:

$$\omega_{aj} = \gamma\tau_{aj}$$

Using this construction, the limiting availability of Expression 12.55 for several cases is shown in Table 12.2.

Similarly, the time-dependent availability function can be obtained using numerical inversion of the Laplace transforms. These functions are shown for the failure replacement, opportunistic failure replacement, and the partial opportunistic age replacement models in Figure 12.6 for the specific case in which $\omega_{a1} = 1.50$, $\omega_{a2} = 2.13$, and $\tau_{a2} = 2.51$, and the life and replacement time distributions are the same as above. The figure serves to illustrate the fact that numerical results are possible. The results also verify the internal consistency of the models.

12.2 NONRENEWAL MODELS

Preventive maintenance, like corrective maintenance, may not imply equipment renewal. For many types of equipment, preventive maintenance reduces failure probability but does not return the system to a “good as new” state. Nevertheless,

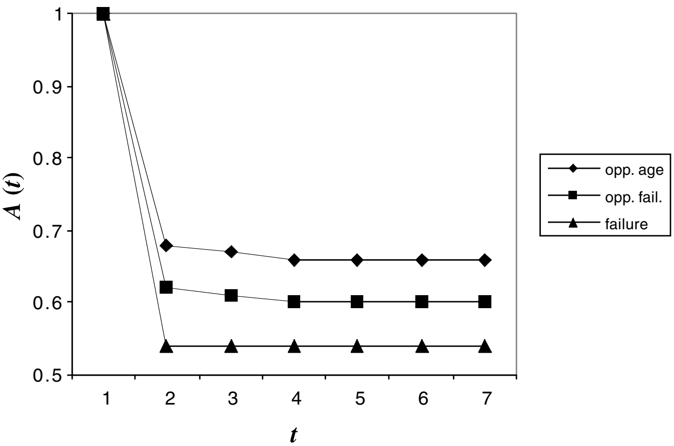


Figure 12.6 Example availability functions. (From Degbotse, A.T. and J.A. Nachlas, "Use of Nested Renewals to Model Availability Under Opportunistic Maintenance Policies," *Proc. of the Annual Reliability and Maintainability Symposium*, pp. 344–350, IEEE, 2003. With permission.)

the reduction in failure probability may make preventive maintenance appealing. Both time- and calendar-based preventive maintenance policies have been defined.

As in the case of the corrective maintenance models of Chapter 10, the portrayal of equipment performance under preventive maintenance depends most heavily on the representation of the postmaintenance failure distribution. Most of the models that have been developed are based on the use of the imperfect repair format for representing the postrepair equipment state. We will examine a few of these models, as well as one based on the Kijima virtual age concept and one based on the quasi-renewal process. Note that, in some of the models, the analysis requires that a finite end to the interval of analysis be defined, while for others, one must assume an infinite time horizon.

One of the earliest and most general models of nonrenewal equipment operation under preventive maintenance is the one defined by Nakagawa [63]. In this model, a block (or periodic) policy is assumed, with minimal repair of failures between PM times and system replacement after N preventive maintenance intervals. Nakagawa treats the post-PM equipment state in the simplest possible manner by assuming only that the system hazard function during the k^{th} interval is $z_k(t)$, where $z_k(t) < z_{k+1}(t)$, so that the system is gradually deteriorating despite the PM schedule. He then assumes all service times are negligible and defines a cost rate function,

$$C(\tau_b, N) = \frac{c_2 \sum_{k=1}^{N} \int_0^{\tau_b} z_k(t) + (N-1)c_1 + c_3}{N\tau_b} \tag{12.73}$$

where c_3 is the cost of system replacement. Note that the assumption of eventual replacement was needed to obtain a tractable model. Note further that N is a decision variable, so the model analysis includes selection of optimal values for both τ_b and N. In fact, Nakagawa shows that the optimal value for τ_b may be obtained by differentiation to be the one for which

$$\sum_{k=1}^{N} \left(\tau_b z_k(\tau_b) - \int_0^{\tau_b} z_k(t)dt \right) = \frac{(N-1)c_1 + c_3}{c_2} \tag{12.74}$$

and the optimal value of N may be determined using the difference expressions:

$$\begin{aligned} C(\tau_b, N) &< C(\tau_b, N-1) \\ C(\tau_b, N) &\le C(\tau_b, N+1) \end{aligned} \tag{12.75}$$

Clearly, the appeal of this model is the fact that the hazard function behavior may be specified arbitrarily (as long as $z_k(t) < z_{k+1}(t)$). The assumption of minimal repair at failure may or may not be viewed as a drawback of the model.

12.2.1 Imperfect PM Models

Clearly, there are numerous possible operating scenarios one might consider. Suppose that PM is imperfect in that a device is renewed with probability p, and the hazard is unchanged by PM with probability $1 - p$. Assume also that a device failure implies device renewal. In this case a periodic (block) PM policy will yield a cost per unit time [63] of

$$C(\tau_b, p) = \frac{\left(c_2 p \sum_{j=1}^{\infty} (1-p)^{j-1} F_T(j\tau_b) + c_1 \sum_{j=1}^{\infty} (1-p)^{j-1} \bar{F}_T(j\tau_b) \right)}{\left(\sum_{j=1}^{\infty} (1-p)^{j-1} \int_{(j-1)\tau_b}^{j\tau_b} \bar{F}_T(u) du \right)} \quad (12.76)$$

because the probability that it is a failure that precipitates renewal is

$$\sum_{j=1}^{\infty} (1-p)^{j-1} \int_{(j-1)\tau_b}^{j\tau_b} f_T(u) du = p \sum_{j=1}^{\infty} (1-p)^{j-1} F_T(j\tau_b)$$

In contrast, if we assume that failure implies minimal repair rather than replacement, then renewal only occurs as a result of (some) PM events. In this case, the cost rate [63] is

$$C(\tau_b, p) = \frac{1}{\tau_b} \left(c_2 p^2 \sum_{j=1}^{\infty} (1-p)^{j-1} \int_0^{j\tau_b} z_T(u) du + c_1 \right) \quad (12.77)$$

where c_2 is the cost of the (minimal) repair that is performed following failure.

In the case of age replacement policies, fewer results have been developed. Block, Borges, and Savits [64] extend their age-dependent imperfect repair model [52] to include PM that is perfect. Let the age-dependent probability of replacement upon failure be denoted by $p(t)$ and the probability of minimal repair upon failure by $q(t) = 1 - p(t)$. Using this format and the associated life distribution of Expression 10.12, Block, Borges, and Savits [64] also allow for the cost of minimal repair to depend upon both time and the number of minimal repairs. They define:

$c_0^k(t)$ = the cost for the k^{th} minimal repair t time units after the most recent device replacement.

Then, recalling that $\mathrm{E}[N_t]$ is the expected number of events during $(0, t)$ in a nonhomogeneous Poisson process that has mean value

$$\int_0^t q(u) z_T(u) du$$

the cost of minimal repairs during an interval $(0, t)$ is $c_0^{E[N_t]+1}(t)$, and the resulting cost per unit time is

$$C(\tau_a, p(t)) = \frac{\left(\begin{array}{l} c_2 F_p(\tau_a) + c_1 \bar{F}_p(\tau_a) + \\ \displaystyle\int_0^{\tau_a} q(x) c_0^{E[N_x]+1}(x) e^{\int_0^x q(u) z_T(u) du} dx \end{array} \right)}{\displaystyle\int_0^{\tau_a} \bar{F}_p(x) dx} \tag{12.78}$$

One appealing feature of this model is that it generalizes several other age replacement imperfect repair models. Specifically, if the minimal repair cost is a constant c_0, the cost model reduces to

$$C(\tau_a, p) = \frac{(c_2 - c_0) F_p(\tau_a) + c_1 \bar{F}_p(\tau_a) + c_0 \displaystyle\int_0^{\tau_a} \bar{F}_p(x) z_T(x) dx}{\displaystyle\int_0^{\tau_a} \bar{F}_p(x) dx} \tag{12.79}$$

and if the probability of minimal repair is constant rather than age dependent, and the minimal repair cost equals the failure replacement cost, the model becomes

$$C(\tau_a, p) = \frac{\dfrac{c_2}{p} F_p(\tau_a) + c_1 \bar{F}_p(\tau_a)}{\displaystyle\int_0^{\tau_a} \bar{F}_p(x) dx} \tag{12.80}$$

where, in this case, $F_p(t)$ is defined by Expression 10.10.

12.2.2 Models Based on the Quasi-Renewal Process

Recall that in a quasi-renewal process [56], the distribution on the k^{th} operating interval, say T_k, is scaled by a constant, α^{k-1}, so that

$$T_k = \alpha T_{k-1} = \alpha^{k-1} T_1$$

and

$$F_{T_k}(t) = \frac{1}{\alpha^{k-1}} F_{T_1}\left(t \middle/ \alpha^{k-1} \right)$$

where $F_{T_1}(t)$ is the underlying life distribution for the device. Recall further that the quasi-renewal function, $M_{F_T}(t)$, is defined in the same manner as a renewal function:

$$M_{F_T}(t) = E[N_t] = \sum_{n=1}^{\infty} F_{S_n}(t)$$

Given these definitions, a block replacement policy (perfect PM) with quasi-renewals upon failure has cost rate

$$C(\tau_b) = \frac{c_2 M_{F_T}(\tau_b) + c_1}{\tau_b} \qquad (12.81)$$

This corresponds directly to the expression for the renewal case and has the corresponding optimality condition.

If the PM that is performed at the scheduled times is imperfect, and failures are quasi-renewal, we obtain a cost rate model of

$$C(\tau_b) = \frac{c_2 p^2 \sum_{j=1}^{\infty} (1-p)^{j-1} M_{F_T}(j\tau_b) + c_1}{\tau_b} \qquad (12.82)$$

Here again, differentiation yields an optimality condition for selecting the value of τ_b.

A rather more interesting model is suggested by Wang and Pham [55]. They consider that, when a new device is placed in operation, it does not undergo PM until after it has experienced some number, say K, of failures. Then, it is subjected to a periodic PM plan. They further consider that the cost of $c_{2f} + (j-1)c_{2v}$ is incurred at the j^{th} failure. Here c_{2f} is the fixed component, and c_{2v} is the variable component of the repair cost. Thus, repair costs are increasing in the number of repairs performed. Wang and Pham also treat both the operating periods and the repair intervals as quasi-renewal processes, so that the length of the j^{th} operating interval has mean $\alpha^{k-1}E[T_1]$, and the elements of the sequence of repair times are scaled by β^{j-1}, so that the duration of the j^{th} repair has mean $\beta^{k-1}E[R_1]$. Next, Wang and Pham assume that, after a device has experienced K failures, it is subjected to imperfect PM at intervals of τ_b and minimal repair at subsequent failures. With this intricate set up, they show that the expected maintenance cost per unit time is

$$C(\tau_b, K, p) = \frac{\left(\begin{aligned} & Kc_{2f} + \frac{K(K-1)}{2} c_{2v} + \frac{c_1}{p} \\ & + pc_{2f} \sum_{j=1}^{\infty} (1-p)^{j-1} \int_0^{j\tau_b} z_T(u/\alpha^k)\,du \end{aligned} \right)}{\frac{1-\alpha^k}{1-\alpha} E[T_1] + \frac{1-\beta^k}{1-\beta} E[R_1] + \frac{\tau_b}{p} + E[R_0]} \tag{12.83}$$

In this expression, the minimal repair cost is taken to be c_{2f}, the time to perform minimal repair is taken to be zero, and the time to perform a perfect repair (renewal) is denoted by R_0. For this policy and model, the corresponding average availability is

$$\frac{\frac{1-\alpha^k}{1-\alpha} E[T_1] + \frac{\tau_b}{p}}{\frac{1-\alpha^k}{1-\alpha} E[T_1] + \frac{1-\beta^k}{1-\beta} E[R_1] + \frac{\tau_b}{p} + E[R_0]} \tag{12.84}$$

If we assume that K is specified on the basis of the equipment design, then we can determine an optimal value of τ_b by differentiation. On the other hand, we can also solve for τ_b as a function of K and then solve for K using finite differences as shown in Expressions 12.75. Finally, note that taking K = 0 or K = 1 in the above two models leads to additional interesting special cases.

12.2.3 Models Based on the Kijima Model

As discussed in Chapter 10, Kijima [53] uses the concept of equivalent or "virtual" age to represent the fact that a device is improved but not renewed by maintenance. The basic expressions that represent this, (10.14 and 10.15), may be represented in general by:

$$A_n = \theta S_n \tag{12.85}$$

where S_n is the actual device age at the time of the n^{th} failure $\left(\sum_{j=1}^{n} T_j\right)$, and A_n is the resulting virtual age just after the completion of the n^{th} repair. As shown in Chapter 10, this construction implies that

$$F_{T_{n+1}}(t) = \Pr[T_{n+1} \le t \mid A_n = u] = \frac{F_T(t+u) - F_T(u)}{\bar{F}_T(u)}$$

The corresponding hazard function during the n + 1st operating interval is

$$\begin{aligned} z_{T_{n+1}}(t) &= z_T(A_n + t - S_n) = z_T(\theta S_n + t - S_n) \\ &= z_T(t - (1-\theta)S_n) \end{aligned} \tag{12.86}$$

and the corresponding hazard at any point in time is

$$\begin{aligned} z_{T_{N_t+1}}(t) &= z_T(A_{N_t} + t - S_{N_t}) = z_T(\theta S_{N_t} + t - S_{N_t}) \\ &= z_T(t - (1-\theta)S_{N_t}) \end{aligned} \tag{12.87}$$

Using this notation, Kijima [53] defines the generalized renewal density as

$$m_{F_A}(t) = E[z_{T_{N_t+1}}(t)] = E[z_T(t-(1-\theta)S_{N_t})] \qquad (12.88)$$

and shows that this generalized renewal density satisfies

$$m_{F_A}(t) = f_T(t) + \int_0^t m_{F_A}(x) \frac{f_T(y-(1-\theta)x)}{\bar{F}_T(\theta x)} dx \qquad (12.89)$$

Clearly, this form corresponds to the Key Renewal Theorem.

Using these definitions, Makis and Jardine [65] examine a modified age replacement policy in which failures prior to τ_a are treated with repairs that improve the device in the sense described by Kijima, and the device is replaced at the first failure to occur after age τ_a. The costs for the repairs are assumed to be time dependent, so the cost per replacement cycle is

$$C(\tau_a) = \frac{c_1 + \int_0^{\tau_a} c_2(u) m_{F_A}(u)du}{E_{F_T}[T] + \int_0^{\tau_a} E_{F_A}[T_{n+1} \mid A_n = \theta u] m_{F_A}(u)du} \qquad (12.90)$$

where the denominator is the expected length of the replacement cycle. Note that the form of the time-dependent repair cost may be selected as appropriate to an application. Makis and Jardine also show that, as long as the repair cost is bounded ($c_2(t) \le K$) and $E_{F_A}[T_{n+1} \mid A_n = y] \ge \varepsilon > 0$, the cost function of Expression 12.90 has a unique minimum that can be computed using the derivative.

One final model is the one defined by Kijima, Morimura, and Suzuki [54] for a block replacement policy with repairs that improve the device age following failure. Since it is assumed that the PM activity involves replacement that is perfect, the model for this policy appears only slightly different from the basic block replacement model. The model is

$$C(\tau_b) = \frac{c_2 M_{F_A}(\tau_b) + c_1}{\tau_b} \qquad (12.91)$$

Clearly, for appropriate choices of the cost parameters and life distribution, this cost function will have a unique minimum. The degree of difficulty for the calculation of the minimum depends mostly on the choice of life distribution. However, Kijima, Morimura, and Suzuki show that, using the now familiar form of the residual life distribution,

$$F_A(t\,|\,x) = \frac{F_T(t+\theta x) - F_T(\theta x)}{\bar{F}_T(\theta x)}$$

the generalized renewal function can be approximated by

$$M_{F_A}(t) \approx \frac{t}{E_{F_T}[T]} - \frac{\int_0^t \bar{F}_T(u)du}{E_{F_T}[T]} + \int_0^t \left(1 - \left\{\frac{\int_0^t \bar{F}_A(u\,|\,x)du}{E_{F_T}[T]}\right\}\left\{\frac{\int_0^{t-x} \bar{F}_A(u\,|\,t-x)du}{\int_0^{\infty} \bar{F}_A(u\,|\,t-x)du}\right\}\right) \left(f_T(x) + \frac{F_T(x)\int_0^x f_A(x-u\,|\,u)du}{\int_0^x \bar{F}_A(x-u\,|\,u)du}\right)dx \qquad (12.92)$$

Depending on the choice of life distribution, this approximation can be very accurate and quite manageable numerically.

12.4 CONCLUSION

Preventive maintenance policies may be defined in several ways and may ultimately have very many operational realizations. The models discussed in this chapter serve to illustrate the many forms the models may take and the two principal approaches to analysis — availability and cost. Preventive maintenance is an essential ingredient in any productivity assurance strategy and for many devices, it is critical to system safety. Given the significance of PM, methods for selecting efficient and effective PM policies are important and are worth the (sometimes taxing) effort they require. It should now be clear that the state of a unit following repair or replacement is the aspect of the equipment behavior that guides the analysis of maintenance policies. It should also be appreciated that numerical approximations to complicated functions will often yield policy solutions that are quite satisfactory.

12.5 ACKNOWLEDGMENT

Material in Section 12.1.4 through Section 12.1.4.5 is based on Degbotse, A.T. and J.A. Nachlas, "Use of Nested Renewals to Model Availability Under Opportunistic Maintenance Policies," *Proc. of the Annual Reliability and Maintainability Symposium*, pp. 344–350, IEEE, 2003, with permission.

12.6 EXERCISES

1. Assume a population of devices display life lengths that are well modeled by a Gamma distribution having parameters $\beta = 3.0$ and $\lambda = 0.005$. Compute and compare the quantities $N(t, \tau_a)$, $N(t, \tau_b)$, N(t), $\tilde{N}(t, \tau_a)$, and $\tilde{N}(t, \tau_b)$ for $t = 100$, 200, and 400 hours and τ_a and τ_b values of 0.1 E[T], 0.2 E[T], and 0.4 E[T].
2. Establish the stochastic ordering between a Weibull distribution and an exponential distribution having the same mean value for the cases of IFR and DFR.
3. Compute an optimal block replacement policy for a system having Gamma life distribution with parameters $\beta = 3.0$ and $\lambda = 0.005$ and assuming the repair cost is 12 times the preventive maintenance cost.
4. Compute an optimal age replacement policy for a system having Weibull life distribution with $\beta = 2.75$ and $\theta = 5000$ hours, where repair cost is 20 times the preventive maintenance cost.
5. Assume a population of devices display life lengths that are well modeled by a Gamma distribution having

parameters β = 2.0 and λ = 0.002. Assume further that the devices are managed using an age replacement policy with τ_b = 750 hours, $G_c(t)$ is exponential with parameter λ = 0.02, and $G_p(t)$ is exponential with parameter λ = 0.05. Compute $A(t)$ for t = 100 and 400 hours, and also $\bar{A}(400, 800)$.

6. Assume a population of devices display life lengths that are well modeled by a Weibull distribution having parameters β = 2.25 and θ = 2000 hours. Assume further that the devices are managed using an age replacement policy with τ_a = 1250 hours, $G_c(t)$ is exponential with parameter λ = 0.008, and $G_p(t)$ is exponential with parameter λ = 0.02. Compute $A(t)$ for t = 1000 and 2400 hours, and also $\bar{A}(2000, 6000)$.
7. Suppose that PM is imperfect in that a device is renewed with probability p, and the hazard is unchanged by PM with probability $1 - p$. Assume also that a device failure implies device renewal, so that the cost rate function is Expression 12.76. Analyze this model as a function of p for a Weibull life distribution having parameters β = 2.25 and θ = 2000 hours, a block replacement policy time of τ_b = 1500 hours, and $c_2 = 20c_1$. Plot the cost rate divided by c_1 as a function of p.
8. Repeat Problem 7 above for the quasi-renewal model of Expression 12.82.
9. Assume a population of devices display life lengths that are well modeled by a Gamma distribution having parameters β = 3.0 and λ = 0.005. Use the approximation for the generalized renewal function of Expression 12.92 to plot the cost rate functions of the Makis and Jardine model of Expression 12.90.
10. Assume a population of devices display life lengths that are well modeled by a Weibull distribution having parameters β = 2.25 and θ = 2000 hours. Use the approximation for the generalized renewal function of Expression 12.92 to plot the cost rate functions of the Kijima et al. model of Expression 12.91.

13

Predictive Maintenance

Increased recognition of magnitude of the potential benefits of preventive maintenance has stimulated an interest in finding new and more efficient PM strategies. One of the most promising new ideas is predictive maintenance (which is also sometimes called condition-based maintenance). The idea of predictive maintenance is that monitoring equipment status should permit the recognition of failure precursors and the corresponding opportunity for preventive intervention.

The development of this idea has followed two principal tracks. One avenue of study has been the investigation of how ordinary process control and process status variables can be interpreted to provide warning of impending failure. For example, one study noted that a rotating cutting tool drew increased quantities of electricity as the tool wore down and approached failure. Another example was an observed increase in the width of the vibration spectrum for a pump as the pump diaphragm deteriorated. As these examples suggest, the focus of this research path is the identification of process variables that provide warnings of failure, the selection of an instrumentation plan that permits monitoring of the variable, and the development of pattern recognition rules that have a high probability of recognizing pathological patterns in the quantities being observed. This is an exciting and interesting domain of study, but it is relatively remote from our probability-based exploration of reliability and maintenance. We will thus not treat this topic further here.

The second primary research direction has been the use of reliability and cost measures to formulate predictive maintenance strategies. Reliability-function-based monitoring plans, policies for inspection, and replacement schedules based both on hazard functions and on cost functions, and adjustment plans for potentially fallible measurements have been addressed. We shall examine some of the models constructed to portray the implications of predictive maintenance.

As with all of the repairable systems analyses we have examined, models that represent predictive maintenance plans differ depending on the assumed state of the system following maintenance. They also differ in the ways in which system deterioration is portrayed. The feature they tend to have in common is the deterioration process and the assumption that there is a measurable variable that represents or is highly correlated with the extent of deterioration. Thus, we start with the representation of system evolution and then move to the analysis of predictive maintenance policies.

13.1 SYSTEM DETERIORATION

We have examined models of component and system failure in Chapter 5. We noted there that the operation of a device usually implies the gradual degradation of the unit and its ultimate failure. Equipment for which this is a reasonable conceptual description are exactly the ones for which predictive maintenance is appropriate. Recognizing this, van Noortwijk, Cooke, and Kok [66] argued that an isotropic deterioration model is appropriate.

The term isotropic is used to represent the idea that, between any two points in time, the deterioration depends on the state of the device at the start of the interval but not on the deterioration prior to the start of the interval. Further, during the interval, deterioration is assumed to proceed in the form of infinitely many very small steps as described by Feller [48] and by Nachlas [30], with the consequence that

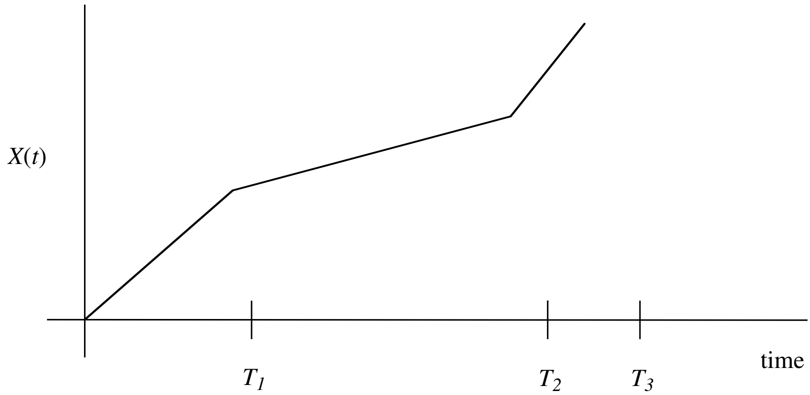


Figure 13.1 Representation of generalized Gamma deterioration process.

deterioration is best modeled using the generalized Gamma process.

Algebraically, we represent the condition of the device by the state variable $X(t)$, for which increments during any interval are random. The magnitude of any increment has a Gamma distribution with a shape parameter that is proportional to the length of the interval. That is,

$$f_{X(t_2)-X(t_1)}(x) = \frac{\lambda^{\beta(t_2-t_1)}}{\Gamma(\beta(t_2-t_1))} x^{\beta(t_2-t_1)-1} e^{-\lambda x} \tag{13.1}$$

Using this model, the average deterioration during the interval is $\beta(t_2 - t_1)/\lambda$, and the variance is $\beta(t_2 - t_1)/\lambda^2$. Thus, there is considerable flexibility in the model, so many types of deterioration can be represented this way. Van Noortwijk, Cooke, and Kok [66] argue that this model provides an accurate representation of cumulative damage processes, erosion and corrosion, defect-based degradation, and most other evolutionary deterioration behaviors. A graphical representation of this type of process is shown in Figure 13.1. Note that the assumption of independent increments implies that the process is semiregenerative, and the times T_1, T_2, and T_3 are semiregeneration points.

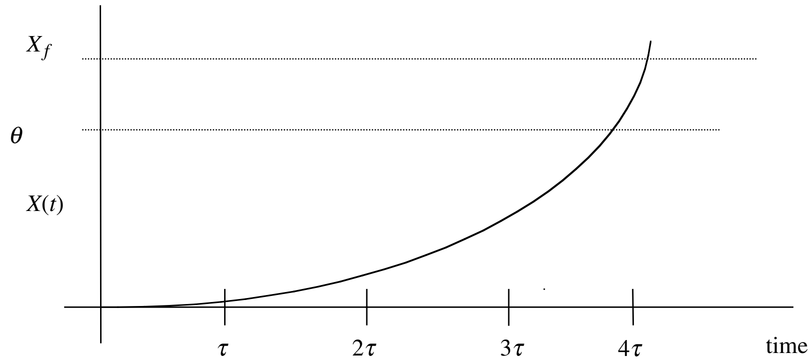


Figure 13.2 Inspection intervals with PM criterion.

13.2 INSPECTION SCHEDULING

The simplest of the predictive maintenance analyses is the definition of a model for selecting inspection schedules. The idea is that a system is gradually deteriorating, but the state of the system may only be determined by inspection, even if it has failed. The problem is to select an inspection schedule and a replacement criterion that will permit us to monitor the state of the system and to renew it prior to failure.

This elementary condition-based replacement strategy for PM is defined like an age replacement policy. Device status is determined by inspections that are equally spaced in time at intervals of length τ. Replacement is performed whenever inspection shows device deterioration to be beyond a PM criterion, say θ, and below a failure threshold, say X_f, and also following failure. This is illustrated in Figure 13.2. For equally spaced inspections and a Gamma deterioration process, the mean deterioration per inspection interval is a linear function of the length of the interval and is $\beta\tau/\lambda$, so the number of intervals until the PM threshold is crossed is

$$\frac{\theta}{\beta\tau/\lambda} = \theta\lambda/\beta\tau \tag{13.2}$$

and the number of inspections required to discover that the threshold has been passed is one greater than that number. Assuming an inspection cost of c_I, and PM and corrective replacement costs of c_1 and c_2, as in Chapter 12, the cost per unit time of an inspection policy τ is

$$C(\tau) = \frac{c_I\left(\theta\lambda/_{\beta\tau} + 1\right) + c_1 F_X(X_f - \theta) + c_2 \bar{F}_X(X_f - \theta)}{\left(\theta\lambda/_{\beta\tau} + 1\right)\tau} \tag{13.3}$$

where $F_X(X_f - \theta)$ is the probability that the deterioration in the final inspection interval does not exceed the difference between the failure threshold and the PM threshold.

Clearly, the simple model of Expression 13.3 is very similar to the age replacement model. One important difference is that analysis of the model permits optimization of both the inspection interval and the PM criterion. A shortcoming of the model is the restriction to equally spaced inspection intervals. A key feature of condition-based maintenance is sensitivity to device state and a corresponding capability to adjust inspection frequency.

13.3 MORE COMPLETE POLICY ANALYSIS

Recently, Grall, Dieulle, Berenguer, and Roussignol [67] defined a model that allows for variable inspection intervals and can be optimized with respect to the replacement threshold. As in the model above, they assume that there is a failure threshold X_f that is fixed by the physical nature of the device. The device is replaced correctively if its state variable passes that threshold and preventively if any inspection shows device state to equal or exceed the PM threshold, θ. In addition, Grall, Dieulle, Berenguer, and Roussignol assume that, following each inspection, the time of the next inspection is selected as a function of device state.

The analysis of a device that is managed adaptively and replaced as described can be based on cost per renewal cycle. However, this can be quite complicated, so instead Grall, Dieulle, Berenguer, and Roussignol based their analysis on the limiting expected cost rate. That is, they formulate the basic cost function as

$$C(t) = c_I N_I(t) + c_1 N_1(t) + c_2 N_2(t) + c_d D(t) \tag{13.4}$$

and then construct

$$E[C_\infty] = \lim_{t\to\infty}\left[\frac{E[C(t)]}{t}\right]$$

$$\begin{aligned} &= c_I \lim_{t\to\infty}\left[\frac{E[N_I(t)]}{t}\right] + c_1 \lim_{t\to\infty}\left[\frac{E[N_1(t)]}{t}\right] \\ &\quad + c_2 \lim_{t\to\infty}\left[\frac{E[N_2(t)]}{t}\right] + c_d \lim_{t\to\infty}\left[\frac{E[D(t)]}{t}\right] \end{aligned} \tag{13.5}$$

In this formulation, $D(t)$ is the quantity of device downtime over $(0, t)$ and c_d is the unit cost of downtime. Also, $N_I(t)$ is the number of inspections, $N_1(t)$ is the number of PM replacements, and $N_2(t)$ is the number of corrective (failure-precipitated) replacements over $(0, t)$.

For mathematical convenience, they assume a minimum inspection interval and then further assume an inspection scheduling decision policy that the time to the next inspection is a linearly decreasing function of system state. That is, following the n^{th} inspection at which the system state is determined to be $X(\tau_n)$, the next inspection time is selected as

$$\tau_{n+1} = \tau_n + 1 + \max\left\{0, \left(\tau_0 - \frac{\tau_0}{\theta - \varepsilon} X(\tau_n)\right)\right\} \tag{13.6}$$

Using this form, the minimum inspection interval is one time unit, and since $X(\tau_0) = 0$, the maximum interval is $1 + \tau_0$ time units. The minimum interval is used whenever the device state is within ε units of the PM threshold, and over time, inspection interval length declines from τ_0 to 1.

As a matter of notational convenience, let $I_n(X(\tau_n))$ represent the time interval until the next inspection that is selected using Expression 13.6. That is

$$I_n(X(\tau_n)) = \tau_{n+1} - \tau_n$$

The Gamma distribution that describes the independent increments in the state variable over any interval, as defined in Expression 13.1, may now be specifically stated as

$$f_{X(I)}(x) = f_{X(\tau_{n+1})-X(\tau_n)}(x) = \frac{\lambda^{\beta I_n(\tau_n)}}{\Gamma(\beta I_n(\tau_n))} x^{\beta I_n(\tau_n)-1} e^{-\lambda x} \quad (13.7)$$

which is to say that the shape parameter of the distribution depends upon the selected length of the inspection interval.

Given the definition of the replacement, inspection, and deterioration processes, the device state, $X(t)$, is a semiregenerative process with regeneration times equal to the times of device replacement. In addition, the discrete time process, Y_n, corresponding to the device state after each inspection, is defined by

$$Y_n = X(\tau_n)$$

and takes real values in $[0, \theta)$. The evolution of the variable Y_n is a continuous state space Markov chain. If we let $\Pi(Y)$ represent the stationary distribution on the state variable Y_n, then the limits in Expression 13.5 can be shown [60] to equal the corresponding first interval expectations with respect to $\Pi(Y)$. That is,

$$\begin{aligned} E[C_\infty] = {} & c_I\left[\frac{E_\Pi[N_I(\tau_1)]}{E_\Pi[\tau_1]}\right] + c_1\left[\frac{E_\Pi[N_1(\tau_1)]}{E_\Pi[\tau_1]}\right] \\ & + c_2\left[\frac{E_\Pi[N_2(\tau_1)]}{E_\Pi[\tau_1]}\right] + c_d\left[\frac{E_\Pi[D(\tau_1)]}{E_\Pi[\tau_1]}\right] \end{aligned} \quad (13.8)$$

Thus, the first step in analyzing the PM strategy is the construction of $\Pi(Y)$.

For the Gamma deterioration process described by the distribution of Expression 13.7, the stationary distribution is obtained as the solution to

$$\Pi(Y) = \int_0^\theta \left(\bar{F}_{X(I)}(\theta - x)\delta_0(y) + y f_{X(I)}(y - x)\right)\Pi(x)dx \quad (13.9)$$

in which $\delta_0(y)$ is the Dirac mass function. The solution of this equation takes the form of the convex combination:

$$\Pi(x) = a\delta_0(x) + (1-a)xg(x) \quad (13.10)$$

where the parameter a $(0 < a < 1)$ must be computed. To compute a, we start with the convolutions

$$\begin{aligned} f_{X(I)}^{(1)}(y-x) &= f_{X(I)}(y-x) = f_{X(\tau_{n+1})-X(\tau_n)}(y-x) \\ f_{X(I)}^{(2)}(y-x) &= \int_x^y f_{X(\tau_{n+1})-X(\tau_n)}(u-x) f_{X(\tau_{n+2})-X(\tau_{n+1})}(y-u)du \\ f_{X(I)}^{(3)}(y-x) &= \int_x^y \int_x^u f_{X(\tau_{n+1})-X(\tau_n)}(w-x) f_{X(\tau_{n+2})-X(\tau_{n+1})} \\ &\quad (u-w) f_{X(\tau_{n+3})-X(\tau_{n+2})}(y-u)dwdu \end{aligned} \quad (13.11)$$

and so on. Then we solve the expression:

$$h(y) = f_{X(I)}(y-0) + \int_0^\theta \left(\sum_{i=1}^{\infty} f_{X(I)}^{(i)}(y-x) \right) f_{X(I)}(x)dx \quad (13.12)$$

and use the result to compute a as

$$a = \left(1 + \int_0^\theta h(y)dy\right)^{-1} \quad (13.13)$$

Once the value of a is determined, the function $g(x)$ in Expression 13.10 is computed as

$$g(x) = \frac{a}{1-a}h(x) \quad (13.14)$$

and the stationary distribution is then well defined.

Once the stationary distribution of the Markov process is obtained, the components of the limiting expected cost rate can be calculated. Specifically,

$$E_{\Pi}[N_1(\tau_1)] = \int_0^{\theta} \left(\bar{F}_{X(I)}(\theta - x) - \bar{F}_{X(I)}(X_f - x)\right)\Pi(x)dx \quad (13.15)$$

$$E_{\Pi}[N_2(\tau_1)] = \int_0^{\theta} \bar{F}_{X(I)}(X_f - x)\Pi(x)dx \quad (13.16)$$

$$E_{\Pi}[D(\tau_1)] = \int_0^{\theta}\left(\int_0^{I(x)} \bar{F}_{X(I(u))}(X_f - u)du\right)\Pi(x)dx \quad (13.17)$$

and for a semigeneration interval corresponding to the time between inspections, $E_{\Pi}[N_I(\tau_1)] = 1$. Finally, the average first cycle length is

$$E_{\Pi}[\tau_1] = \int_0^{\theta} I(x)\Pi(x)dx \quad (13.18)$$

The evaluation of these expressions appears to require considerable numerical effort but actually, Grall, Dieulle, Berenguer, and Roussignol [67] show that the computational effort is quite manageable. Taking $\theta = 6$ to be fixed and $X_f = 12$ with costs of $c_I = 25$, $c_1 = 50$, $c_2 = 100$, and $c_d = 250$ and distribution parameters of $\beta = \lambda = 1.0$, they compute the optimal values of $\tau_0 = 6.5$ and $\varepsilon = -3$ which define the inspection policy. (The negative value of ε means the minimum feasible interval duration is not used.) In a separate analysis, they assume $X_f = 60$ with costs of $c_I = 2$, $c_1 = 90$, $c_2 = 100$, and $c_d = 100$ and distribution parameters of $\beta = 1.0$ and $\lambda = 0.2$ and compute the optimal policy parameters to be $\theta = 50$, $\tau_0 = 5.4$, and $\varepsilon = 5$. In general, they illustrate the fact that the numerical integration necessary to evaluate the model can be accomplished efficiently.

13.4 CONCLUSION

The concept of predictive maintenance is very appealing. Predictive maintenance holds the promise of significant maintenance cost reductions and associated increases in productivity. The definition of alternate policy formats and the formulation and analysis of models for selecting predictive maintenance policies is embryonic and represents one of the important new frontiers in maintenance planning research.

13.5 EXERCISES

1. Assume that the Gamma process of Expression 13.1 has parameters $\beta = 3.0$ and $\lambda = 0.5$, and simulate the process over an interval of 20 time units with inspections every 4 time units. Plot the observed values of $X(t)$ relative to a failure PM threshold of $\theta = 100$.
2. Assume a Gamma deterioration process has parameters $\beta = 2.0$ and $\lambda = 0.4$ and cost parameters of $c_I = 2$, $c_1 = 10$, and $c_2 = 20$. Solve the model of Expression 13.3 to determine the optimal inspection schedule.
3. Assume a device has a Gamma deterioration process having parameters $\beta = 5.0$ and $\lambda = 1.0$, failure and PM thresholds of $X_f = 325$ and $\theta = 300$, and an inspection schedule that is defined by $\tau_0 = 8$ and $\varepsilon = 20$. Use Expressions 13.10 through 13.14 to determine the stationary distribution on the state of the device.
4. For the device described in Problem 3, assume cost parameters of $c_I = 5$, $c_1 = 25$, $c_2 = 75$, and $c_d = 200$, and calculate the limiting expected cost rate for predictive maintenance of the device.

14

Special Topics

To conclude our study of reliability and maintenance planning models, we should examine some special topics that build upon the topics treated in this text. These are not the only extensions to the material in this text, but they are reasonably general and relatively close to the preceding discussions. The first of these extensions is warranties. Following a discussion of warranty policies, the idea of reliability growth is examined. Then we consider reliability models for dependent components, and finally, we explore the construction and use of bivariate and multivariate reliability models.

14.1 WARRANTIES

A warranty is a guarantee by a producer that a product will display a defined level of reliability. It is increasingly common for manufacturers to provide warranties for their products. The chief reason that manufacturers provide warranties is that customers often consider warranties to be appropriate, and warranties are thus an important ingredient in successful marketing. Independent of this point, it is generally recognized that complex and especially expensive products should be guaranteed to function properly. In 1964, the U.S. Congress enacted the Magnuson Moss Act, which formally defined how warranties are to be structured and specified the responsibilities of manufacturers in meeting warranty commitments.

There are essentially two types of warranties. These are full replacement warranties and pro rata warranties. Pro rata warranties are usually offered on products such as automobile components (e.g., tires and batteries). Common characteristics of products that carry pro rata warranties are that (a) product use implies wear or at least accumulating deterioration, (b) repair is either physically impractical or economically inefficient, and (c) evaluation of product age is reasonably straightforward. Under a pro rata warranty, the manufacturer returns a proportion of the original price of the product to the customer in the event of product failure. The proportion of the price that is returned is computed on the basis of an estimate of how much of the product life the customer has lost due to failure of the product. In the case of an automobile tire, a blowout renders the tire unusable. If one occurs, it is common to measure the depth of the tire tread that remains and to use the ratio of the remaining tread depth to the original tread depth to determine the tire life lost due to the failure. The formula for converting the lost life into a cash settlement is usually specific to the manufacturer.

The full replacement warranty is a guarantee to repair or replace a failed product (or component of a product) in order that the product be as good as new. A key feature of the full replacement warranty is the definition of the time limit on its applicability. For some products, that warranty is in force for a fixed period of time from the date of product purchase. A three-year warranty on an automobile or a one-year warranty on a stereo are examples. In the case of the auto, the frequency of repair does not change the termination date of the warranty.

For some products, the full replacement warranty is renewed when the product is replaced under the warranty agreement. For example, a portable stereo that fails prior to the completion of its one-year warranty period might be replaced with another copy of the product that is also warranted for one year. Most durable consumer goods carry one of the two types of warranties.

The choice of which type of warranty to offer is made by the manufacturer and is highly influenced by the nature of the product and its reliability. Questions about warranty policies for which analysis is meaningful include the length of the warranty period, the cost that is expected to accrue in meeting warranty commitments, and the formula to apply for the pro rata case. We shall examine some simple models for the warranties below. For each model, we assume that the warranty promotes demand and therefore increases profit.

14.1.1 Full Replacement Warranties

Most analysts distinguish between repairable and nonrepairable products. From an analytical perspective, this distinction is unnecessary if the product is as good as new after warranty service. Assume that the warranty for a product is renewed at a repair time, and let T_w represent the duration of the full replacement warranty. Now there are many scenarios that one might assume for the warranty. For now, we suppose that the cost of providing warranty support for a single unit of product is the cost of manufacturing the product plus an operating or processing cost that is proportional to the selling price of the product. That is, the cost to provide a warranted replacement is $(C + \delta P)$, where C is the production cost for the unit of product, P is its selling price, and δ is the proportionality factor that represents the fraction of the selling price consumed in the processing of a warranty claim. Assume further that we can represent the product generation cost as a fraction of the selling price:

$$C = \gamma P$$

Next, assume that that the likelihood of product sales is enhanced by the warranty, and that this can be modeled using

$$\mathrm{E}[\text{demand} \mid T_w] = d(T_w) = \left(u_1 + u_2 T_w^r\right)$$

where $0 \le r < 1$. The expected profit realized from product sales adjusted by the cost of warranty support is

$$E[\mathit{profit}] = (P - C)d(T_w) - E[N(T_w)](C + \delta P) \qquad (14.1)$$

where

$$E[N(T_w)] = \frac{1}{\bar{F}_T(T_w)} - 1 \tag{14.2}$$

is the expected number of replacement units provided each purchaser. The reason this is the number of copies used to meet each purchase is that the number of copies of the product used in order to achieve a life duration in excess of the warranty interval is geometric, with parameter $\bar{F}_T(T_w)$, the survivor function for the product life distribution. The expected value of the number of replacements, $N(T_w)$, is one less than the number of copies used. Substituting the demand and cost expressions along with $E[N(T_w)]$ into Expression 14.1 yields the expected profit model:

$$E[profit] = (P - \gamma P)\left(u_1 + u_2 T_w^r\right) - \left(\frac{1}{\bar{F}_T(T_w)} - 1\right)(\gamma P + \delta P)$$

$$= (1-\gamma)Pu_1 + (1-\gamma)Pu_2 T_w^r - \frac{(\gamma+\delta)P}{\bar{F}_T(T_w)} + (\gamma+\delta)P \tag{14.3}$$

Taking the derivative of Expression 14.3 yields

$$\frac{d}{dT_w} E[profit] = r(1-\gamma)Pu_2 T_w^{r-1} - \frac{(\gamma+\delta)Pf_T(T_w)}{\left(\bar{F}_T(T_w)\right)^2}$$

and a convenient application of the hazard function identity makes this

$$\frac{d}{dT_w} E[profit] = r(1-\gamma)Pu_2 T_w^{r-1} - \frac{(\gamma+\delta)Pz_T(T_w)}{\bar{F}_T(T_w)} \tag{14.4}$$

One reason this is convenient is that the second derivative expression is

$$\frac{d^2}{dT_w^2}E[\mathit{profit}] = r(r-1)(1-\gamma)Pu_2T_w^{r-2}$$

$$-\frac{(\gamma+\delta)Pz_T(T_w)f_T(T_w)}{\left(\bar{F}_T(T_w)\right)^2} - \frac{(\gamma+\delta)P}{\bar{F}_T(T_w)}\frac{d}{dT_w}z_T(T_w)$$

and clearly this is negative for any IFR distribution. For other distributions it should be checked but will usually be negative. Thus, the solution of Expression 14.4 will usually represent a profit-maximizing warranty policy time. That solution is obtained numerically as the value of T_w for which

$$\frac{z_T(T_w)}{T_w^{r-1}\bar{F}_T(T_w)} = \frac{r(1-\gamma)u_2}{(\gamma+\delta)} \tag{14.5}$$

As an example, if we assume that production cost is 75% of selling price for a product, and that δ = 0.025, u_2 = 100, r = 0.1, and the device has a Weibull life distribution with β = 2.75 and θ = 4 months, the computed warranty time is 3.5 months.

Note that the above formulation yields a solution that is independent of the price of the product and depends instead upon the proportion of revenue returned to the customer and upon the failure behavior of the product.

Next consider the situation in which the warranty has a fixed duration, during which replacement (repair) service is provided but the duration of the warranty is not extended. For this case, the cost of meeting the warranty commitment for each unit of product is defined by the renewal function based on the product life distribution function. If it is again assumed that demand depends on the duration of the warranty, as expressed by $d(T_w)$, the expected profit function is

$$E[\mathit{profit}] = (P-\gamma P)\left(u_1+u_2T_w^r\right) - M_{F_T}(T_w)(\gamma P+\delta P)$$

$$= (1-\gamma)Pu_1 + (1-\gamma)Pu_2T_w^r - M_{F_T}(T_w)(\gamma+\delta)P \tag{14.6}$$

To identify the optimal warranty policy time, we again use the derivative

$$\frac{d}{dT_w}E[profit] = r(1-\gamma)Pu_2T_w^{r-1} - (\gamma+\delta)Pm_{F_T}(T_w) \quad (14.7)$$

so the optimal policy time is the value of T_w for which

$$\frac{m_F(T_w)}{T_w^{r-1}} = \frac{r(1-\gamma)u_2}{(\gamma+\delta)} \quad (14.8)$$

For the example values used above, numerical solution yields an optimal warranty policy time of 4.2 months.

14.1.2 Pro Rata Warranties

The chief difference between a full replacement and a pro rata warranty is that the resources committed to meeting the warranty vary for the pro rata case. Under a full replacement warranty, the producer incurs the full cost of product replacement when a warranted item fails. While this may be a fraction of the sale price of the product, it is still constant over the duration of the warranty. In the case of a pro rata warranty, the producer usually "pays" the customer a proportion of the price of the warranted product, and the proportion paid depends on the condition or age of the product when it fails.

To represent the effect of pro rata warranty policy on profitability, define a model that describes the costs and profits associated with a single unit of product. One approach is to consider that the warranty affects the likelihood of repeat purchase and that the original purchase may be assumed. That is, we might assume that for each unit of product sold, the probability of a repeat purchase following failure of the unit is proportional to the ratio of failure age to policy age, say:

$$\text{E}[\text{demand} \mid T_w] = d(T_w) = \eta\left(\frac{t}{T_w}\right)^s$$

Now, assume as before that the consequences of failure are warranty costs of $(C+\delta P)$ and that the revenue resulting from the repurchase is $(P-C)=(1-\gamma)P$. The dependence of warranty cost on product age may be modeled by making the proportion a function of time. A plausible general form for this function is

$$\delta(t)=\delta_0\left(1-\left(t/T_w\right)^r\right) \tag{14.9}$$

This form permits either linear or nonlinear reduction in warranty value and also allows that the maximum warranty value be any fraction of the purchase price. Using this form, the profit associated with the warranty assigned to each unit of product is

$$E[profit]=P\int_0^{T_w}(1-\gamma)\eta\left(\frac{t}{T_w}\right)^s f_T(t)-\delta_0\left(1-\left(t/T_w\right)^r\right)f_T(t)dt$$

$$=P\int_0^{T_w}\left((1-\gamma)\eta\left(\frac{t}{T_w}\right)^s-\delta_0\left(1-\left(t/T_w\right)^r\right)\right)f_T(t)dt \tag{14.10}$$

The first derivative condition for this function is

$$\frac{d}{dT_w}E[profit]=P(1-\gamma)\eta f_T(T_w)$$

$$-P\int_0^{T_w}\left(\frac{s(1-\gamma)\eta t^s}{T_w^{s+1}}+\frac{\delta_0 t^r}{T_w^{r+1}}\right)f_T(t)dt=0$$

and the second derivative confirms that the corresponding solution is a maximum. For the numerical values of the previous examples, we add $s = 1.5$ and $\eta = 0.8$. The model solution is a warranty policy time of 5.25 months.

Naturally, there are many possible variations to the models described here. One extension that is often considered worthwhile is to apply a discount factor to future cash flows. Depending upon the application, this and other modifications may be reasonable. The modeling format should be the same, and the result should be an effective approach to the selection of a warranty policy.

14.2 RELIABILITY GROWTH

New products and new product designs are usually enhanced during some sort of "development process." This development process generally includes considerable testing and adjustment of both the design and the manufacturing processes used to create the product. The adjustments are frequently of sufficient significance that they are labeled "engineering design changes." An important element of the enhancement of the product design and manufacturing process formulation is an increase in the reliability of the product. This improvement in product reliability is called reliability growth. It has been observed in many products.

The initial study of reliability growth was reported by Duane [68], and a subsequent enhancement of the "Duane reliability growth model" was provided by Crow [69]. Crow's model is often referred to as the AMSAA (Army Material Systems Analysis Activity) model. The models defined by both Duane and Crow are now incorporated in *Military Handbook 189* [70], which specifies procedures for performing reliability growth evaluation tests. A key point concerning the reliability growth models is that they are based on an assumption that the life distribution is exponential (constant hazard).

The fundamental idea underlying the reliability growth models is that, during development testing, some of the failures that are observed point to failure modes that result from design flaws or manufacturing errors that can be eliminated. Once those problems are eliminated, the product reliability is improved, and the improvement can be described by a decrease in the hazard rate or, equivalently, by an increase in the mean life length for the product.

Having observed the reduction in hazard rate for several product designs, Duane formulated a general power law model for reliability growth. Assuming that a series of device designs is tested over τ units of time and that $n(\tau)$ of the test specimens fail through time τ, the cumulative hazard rate may be estimated by $n(\tau)/\tau$. Duane observed that the reciprocal of $n(\tau)/\tau$ could be called the "cumulative mean time to failure" (MTTF), and that it appears to be well fit by the power law function,

$$\tau/n(\tau) = k\tau^b \tag{14.11}$$

which implies that the logarithm should be well fit by the linear model:

$$\ln\left(\tau/n(\tau)\right) = \ln k + b\ln\tau \tag{14.12}$$

The parameter b is called the rate of growth. Typical observed values for b are in the range of 0.3 to 0.6. In any case, the derivative of the cumulative hazard function yields the hazard rate. For the Duane model,

$$\hat{\lambda}(\tau) = \frac{1-b}{k}\tau^{-b} \tag{14.13}$$

or equivalently,

$$\hat{E}[t] = \frac{1}{\hat{\lambda}(\tau)} = \frac{k}{1-b}\tau^b \tag{14.14}$$

Observe that one implication of this model is that one may set a design reliability target and then compute how long the development testing should be continued in order to attain the target. For any specific design, this implication and the assumption of constant hazard require careful consideration.

The AMSAA model developed by Crow [69] provides three important improvements to the Duane model. First, Crow specifically distinguishes between failures observed during testing that result from failure modes that cannot be altered and those failures that result from design or manufacturing problems that can be addressed. Second, he allows for the possibility that adjustments to the product design or manufacturing process may not eliminate the problem completely. Instead, for each failure mode addressed, Crow includes an "effectiveness factor" in the calculation of the reliability improvement. Finally, Crow defines the growth model for a sequence of test cycles. He assumes that product design or manufacturing process changes are implemented at the end of each of several test cycles. Thus, the development test is described by the set of test cycle durations (τ_1, τ_2, ..., τ_m). Since the life distribution is assumed to be exponential, the number of failures that occur during each test cycle is Poisson distributed with mean equal to $\lambda_j\tau_j$, where λ_j is the sum of the hazard rates for the correctable and the uncorrectable failure modes during cycle j. (The failure modes are all assumed to be mutually independent.) Thus, under the Crow model, the hazard rate during any test cycle is

$$\lambda_j = \lambda_u + \sum_{i=1}^{j-1}\sum_{k=1}^{K}(1-d_{ik})\lambda_{ck} \tag{14.15}$$

where λ_u is the hazard rate associated with the uncorrectable failure modes, λ_{ck} is the hazard rate for the k^{th} correctable failure mode, d_{ik} is the effectiveness in modifying the k^{th} failure mode of the engineering change implemented at the end of cycle i, and K is the number of correctable failure modes.

As in the case of the Duane model, the Crow model is fit to the power law function stated in (14.11). The model is then used to estimate the growth rate, b, and to estimate the final hazard (and mean life) or to set test durations.

14.3 DEPENDENT COMPONENTS

The possibility of dependence among system components is an important consideration in the analysis of system reliability. The dependence may arise as a consequence of sharing of loads, susceptibility to the same failure causes, or mutual interactions. Two examples of mutual interactions are (1) the case of microelectronic components that are located near enough that the heat generated by each component affects the thermal environment experienced by the other component and (2) mechanical components such as gears and struts that transmit forces from one component to another.

Naturally, there are several reasonable approaches to modeling component dependencies. One general approach is to simply formulate a system reliability model for which the system continues to operate but the hazard function changes when a component fails. This model seems most representative of the shared load case. The general model is

$$\bar{F}_{T_s}(t) = \bar{F}_{T_1}(t) + \int_0^t f_{T_1}(u)\bar{F}_{T_2}(t-u)du = e^{-Z_{T_1}(t)} + \int_0^t f_{T_1}(u)e^{-Z_{T_2}(t-u)}du \tag{14.16}$$

where T_s is the overall system life length, T_1 is the time at which the first component failure occurs, and T_2 is the residual life length of the system following the first component failure. If this model is applied to a shared load system, one often assumes that the individual component life lengths are independent, and it is only the sharing of the load that produces the dependence at the system level. In that case,

$$\bar{F}_{T_1}(t) = \bar{F}_{X_1}(t \mid y(T_1))\bar{F}_{X_2}(t \mid y(T_1)) \tag{14.17}$$

where $y(T_1)$ describes the effect of load sharing on the life distributions. Similarly,

$$\int_0^t f_{T_1}(u)\bar{F}_{T_2}(t-u)du = \int_0^t \left(f_{X_1}(u)\bar{F}_{X_2}(u)\frac{\bar{F}_{X_2'}(t-u)}{\bar{F}_{X_2}(u)} + f_{X_2}(u)\bar{F}_{X_1}(u)\frac{\bar{F}_{X_1'}(t-u)}{\bar{F}_{X_1}(u)} \right) du$$

$$= \int_0^t \left(f_{X_1}(u)\bar{F}_{X_2'}(t-u) + f_{X_2}(u)\bar{F}_{X_1'}(t-u) \right) du \tag{14.18}$$

where the notation X_i' is intended to signify the fact that the hazard function for the surviving component is different than when both components are functioning.

An alternate general model for component dependence was proposed by Barlow and Proschan [11]. They argued that a reasonable interpretation of certain shock models leads directly to a bivariate exponential system life distribution. Their construction proceeds as follows. Suppose a system composed of two components (in arbitrary structure) is subject to shocks from three sources. The shocks from each source occur according to Poisson processes, such that $N_i(t)$ is the number of shocks from source i occurring during $(0, t)$, and λ_i is the parameter for shock process i. Suppose further that a shock from source 1 causes failure of component 1 with probability θ_1, while a shock from source 2 causes failure of component 1 with probability θ_2. In the case of shocks from source 3, each shock causes failure of both components with probability θ_{11}, causes failure of component 1 only with probability θ_{10}, causes failure of component 2 only with probability θ_{01}, and results in no failures with probability θ_{00}. Then the joint survival probability for life lengths T_1 and T_2 for the two components is

$$\Pr[T_1 > t_1, T_2 > t_2] = \left(\sum_{k=0}^{\infty} e^{-\lambda_1 t_1} \frac{(\lambda_1 t_1)^k}{k!} (1-\theta_1)^k\right) \left(\sum_{k=0}^{\infty} e^{-\lambda_2 t_2} \frac{(\lambda_2 t_2)^k}{k!} (1-\theta_2)^k\right)$$
$$\left(\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left\{e^{-\lambda_3 t_1} \frac{(\lambda_3 t_1)^j}{j!} \theta_{00}^j\right\} \left\{e^{-\lambda_3 (t_2 - t_1)} \frac{(\lambda_3 (t_2 - t_1))^k}{k!} (\theta_{00} + \theta_{10})^k\right\}\right) \tag{14.19}$$

when $0 \le t_1 \le t_2$. Taking the sums and simplifying, they obtain

$$\Pr[T_1 > t_1, T_2 > t_2] = e^{-t_1(\lambda_1\theta_1 + \lambda_3\theta_{10}) - t_2(\lambda_2\theta_2 + \lambda_3(1-\theta_{00}-\theta_{10}))} \tag{14.20}$$

and by symmetry, when $0 \le t_2 \le t_1$,

$$\Pr[T_1 > t_1, T_2 > t_2] = e^{-t_1(\lambda_1\theta_1 + \lambda_3(1-\theta_{00}-\theta_{01})) - t_2(\lambda_2\theta_2 + \lambda_3\theta_{01})} \quad (14.21)$$

Combining these expressions yields the standard bivariate exponential survivor function:

$$\Pr[T_1 > t_1, T_2 > t_2] = e^{-\lambda_1^* t_1 - \lambda_2^* t_2 - \lambda_3^* \max(t_1, t_2)} \quad (14.22)$$

where $\lambda_1^* = \lambda_1\theta_1 + \lambda_3\theta_{10}$, $\lambda_2^* = \lambda_2\theta_2 + \lambda_3\theta_{01}$, and $\lambda_3^* = \lambda_3\theta_{11}$.

14.4 BIVARIATE RELIABILITY

People frequently discuss equipment behavior in terms of age and usage. Common examples are automobiles and automobile tires in which both model year and accumulated mileage are usually included in discussions of longevity. Less well-recognized examples for which two measurement scales are quite important include factory equipment, power generation machines, and aircraft. In fact, the longevity of many of the devices that reliability specialists study is meaningfully described in terms of two measures.

Device life is a resource that may be best represented and for which the consumption may best be measured using a two (or higher) dimensional vector, and the quantities that comprise the vector are specific to the equipment. Years of usage and mileage are not the only two quantities that might describe device longevity. We use the terms age and use here, but these terms are generic and may represent quite different measures than duration of ownership and distance traveled. In the example of an automobile tire, age might correspond to accumulated mileage and usage might be measured as tread loss. Even more complicated measures such as current flow and thermal history may be appropriate for some integrated circuits.

There are basically two ways to approach the definition of a bivariate reliability model. The traditional approach has been to define the second variable, usage, as a function of time, so that the bivariate model can be collapsed into a univariate model. Models of this type are discussed first. In fact, we have already discussed some, as the cumulative damage models portray equipment reliability in terms of deterministically defined deterioration occurring at random points in time, and the proportional hazards models treat age as a deterministic function of use covariates. Clearly, the nondeterministic cases are more interesting.

A second approach that is more recent and was first treated by Singpurwalla and Wilson [71] and subsequently by Yang and Nachlas [72] is to develop reliability functions that are truly bivariate. These are presented subsequently.

14.4.1 Collapsible Models

The collapsible models are generally defined so that the distribution on "usage" is a stochastic function of age. The cumulative damage models and particularly the shot noise process discussed in Chapter 5 and the Gamma process models treated in Chapter 13 are examples of the type of formulation.

The basic formulation is to define age and usage variables for which the distribution on usage depends upon device age. From a statistical perspective, this is a reasonable concept. For example, it would be appropriate to describe a distribution on the age (or wear) for a population of three-year-old cars.

Algebraically, the collapsible models are quite complicated. The most direct formulation is obtained by defining the usage variable, say X, as a stochastic process and defining the failure to correspond to values of the state variable that exceed a threshold, say θ (as in Chapter 13). Then for the bivariate survival function, $\bar{F}_{T,X}(t,x)$, the time evolution is described by

$$\frac{\partial}{\partial t}\bar{F}_{T,X}(t,x) = -\left(\lambda + z_{T,X)}(t,x)\right)\bar{F}_{T,X}(t,x) + \lambda\int_0^x \bar{F}_{T,X}(t,u) f_{X(t)}(\theta - u)\,du \tag{14.23}$$

where $f_{X(t)}(x)$ is the density function for the distribution on the transition in state over time.

In Chapter 13, we treated state transitions as a Gamma process. This is often an appropriate choice. If the state transition process is to be Gaussian, the result is a Weiner (rather than Gamma) process, and this model may be representative of some devices, particularly those that experience healing or attenuation of damage.

In any case, the general model of Expression 14.23 permits very many specific cases. The following is one of the more manageable possibilities. If the hazard function is assumed to be an additive function of independent time and usage processes so that

$$z_{T,X}(t, x) = z_T(t) + \eta x \tag{14.24}$$

then the Laplace transform of Expression 14.23 with respect to the state variable X is

$$\bar{F}^*_{T,X}(t, s) = \exp\left(\frac{\lambda}{\eta}\int_s^{s+\eta t} f^*_{X(t)}(v)dv - \lambda t - \int_0^t z_T(v)dv\right) \tag{14.25}$$

Then, taking the state transition process to be a Gamma process, the collapsed time dependent life distribution is obtained as the Laplace transform evaluated at $s = 0$:

$$\bar{F}_T(t) = \bar{F}^*_{T,X}(t, 0) = \exp\left(\frac{\lambda}{\eta}\int_0^{\eta t} f^*_{X(t)}(v)dv - \lambda t - \int_0^t z_T(v)dv\right) \tag{14.26}$$

Nearly all other reasonable choices of constituents of the above model are more difficult to evaluate, but most that are practically interesting can be managed numerically.

14.4.2 Bivariate Models

There are actually two classes of bivariate models that are not collapsed into a single dimensional form. These are the models based on a stochastic functional relationship between the two variables and the models that represent the variables as correlated.

14.4.2.1 Stochastic Functions

The definition of failure models on the basis of stochastic functions relating age and use starts with the specification of how the stochastic feature of the longevity variables is portrayed. Assume that the time and use to failure are related by the function $u = B(t)$, and that the stochastic nature of this relationship can be represented by treating one or more of the parameters of $B(t)$ as random variables.

The interpretation of the function $B(t)$ is that, across a population of devices, the accumulated usage by age t is $B(t)$. This is equivalent to saying that the mileage traveled by two-year-old cars is $B(t = 2)$. Of course, we impose a probability distribution on $B(t)$ to model its dispersion. To illustrate this construction, we consider four example forms here:

1. $B(t) = \alpha t + \beta$
2. $B(t) = \alpha t^2 + \beta t + \gamma$
3. $B(t) = \alpha t^n$
4. $B(t) = (e^{\alpha t} - 1)/(e^{\alpha t} + \beta)$

where the fourth form is the logistic model analyzed by Eliashberg, Singpurwalla, and Wilson [73]. In each case, we introduce randomness into the function by treating the parameter α as a random variable having distribution $\pi_\alpha(\cdot)$. This imposes random variation on the extent of use experienced by any age. Consequently, both age and usage at failure are random variables. Certainly, there may be many other functional forms that may be defined and that may be representative of observed behavior. The analytical methods described here may apply to those other forms as well.

The use of the distribution $\pi_\alpha(\cdot)$ to construct the marginal probability distribution on usage is accomplished using well-known methods. In general, as indicated by Eliashberg, Singpurwalla, and Wilson [73], we may construct the marginal density on use as

$$f_U(u) = f_{g(t)}(u) = \left| \frac{d\alpha(u)}{du} \right| \pi_\alpha(\alpha(u)) \tag{14.27}$$

For example, with $B(t) = \alpha t + \beta$, solving for α yields

$$\alpha(u) = \frac{u - \beta}{t} \quad \text{and} \quad \frac{d\alpha(u)}{du} = \frac{1}{t}$$

so

$$f_U(u) = \frac{1}{t}\pi_\alpha\left(\frac{u - \beta}{t}\right) \tag{14.28}$$

Once the marginal distribution on usage is obtained, we construct the joint failure density using the well-known conditioning relation:

$$f_{T,U}(t, u) = f_{T|U}(t)\, f_U(u) \tag{14.29}$$

and the conditional density $f_{T|U}(t)$ is obtained using the well known relationship between a density and its hazard function:

$$f_{T|U}(t) = z_{T|U}(t) = z_{T|g(t)}(t)\exp\left\{-\int_0^t z_{T|g(t)}(x)dx\right\} \tag{14.30}$$

We use this form specifically so that we can focus upon the hazard function in the definition of the failure model. We assume that the bivariate device failure hazard function may be stated as

$$z_{T,U}(t, u) = \lambda(t) + \eta(u) \tag{14.31}$$

so that the definitions of the functions $\lambda(t)$, $\eta(u)$, and $B(t)$ determine the hazard and ultimately the bivariate life distribution. Here, we treat the simplest conceivable form of the hazard function. More intricate, and perhaps realistic, forms should be studied. Thus, we assume that $\lambda(t)$ and $\eta(u)$ are simple linear functions. That is, we use $\lambda(t) = \lambda t$ and $\eta(u) = \eta u$.

Under this modeling format, the bivariate life distribution corresponding to form (1) above is obtained by constructing:

$$z_{T|U}(x) = z_{T|g(t)}(x) = \lambda x + \eta\left(\left(\frac{u - \beta}{t}\right)x + \beta\right) \tag{14.32}$$

and applying Expressions 14.29 and 14.30 to obtain

$$f_{T,U}(t,u) = \frac{\lambda t + \eta u}{t} \exp\left\{-\frac{\eta(u+\beta)}{2}t - \frac{\lambda}{2}t^2\right\}\pi_\alpha\left(\frac{u-\beta}{t}\right) \quad (14.33)$$

The same analytical approach yields

$$f_{T,U}(t,u) = \frac{\lambda t + \eta u}{t^2} \exp\left\{-\frac{\eta(u+2\gamma}{3}t\right\}$$

$$\exp\left\{-\frac{3\lambda + \eta\beta}{6}t^2\right\}\pi_\alpha\left(\frac{u-\beta t-\gamma}{t^2}\right) \quad (14.34)$$

$$f_{T,U}(t,u) = \frac{\lambda t + \eta u}{t^n} \exp\left\{-\frac{\eta u}{n+1}t - \frac{\lambda}{2}t^2\right\}\pi_\alpha\left(\frac{u}{t^n}\right) \quad (14.35)$$

and

$$f_{T,U}(t,u) = \frac{(1+\beta)(\lambda t + \eta u)}{t(1-u)(1+\beta u)}$$

$$\pi_\alpha\left(\frac{1}{t}\ln\frac{1+\beta u}{1-u}\right) \exp\left\{\frac{\eta}{\beta}t - \frac{\lambda}{2}t^2 - \frac{\eta\frac{\beta+1}{\beta}}{\ln\frac{1+\beta u}{1-u}}\ln\frac{1}{(1-u)}t\right\} \quad (14.36)$$

for cases (2), (3), and (4), respectively. Note that in case (4), the definition of the use function limits the variable U to [0,1], so the functions may require rescaling for some applications. Also, in cases (1) and (2), the forms of the functions $B(t)$ allow a nonzero minimum value for usage.

Finally, observe further that all four models are well defined and require only the specification of the density $\pi_\alpha(\cdot)$ to be complete bivariate life distributions. On the other hand, for each of them, it is unlikely that a closed-form expression can be obtained for the marginal distribution on age at failure. In any case, the above examples illustrate the construction of a model in which the usage variable is a stochastic function of the age variable.

14.4.2.2 Correlation Models

In many applications, the two life variables appear to be correlated rather than functionally dependent. The definition of models that can represent correlations in the life variables appears initially to be somewhat simpler than the construction above. We simply choose a bivariate distribution. However, it is important that the distribution be capable of accurately representing equipment behavior, and in particular, that it have marginal distributions that are consistent with experience. Three example models that appear to hold promise for representing bivariate failure processes in which the two variables are correlated are described here. Once again, our choices are not the only conceivable ones. Observed equipment behavior may suggest the use of a different distribution, and the ideas developed here should apply to those cases as well as the ones treated here.

The first of the candidate models is the generalization of the bivariate exponential model defined by Baggs and Nagagaja [74]. In this model, the reliability function is

$$\overline{F}_{T,U}(t, u) = e^{-(\lambda t+\eta u)}\left(1+\rho(1-e^{-\lambda t})(1-e^{-\eta u})\right) \qquad (14.37)$$

so the corresponding density function is

$$f_{T,U}(t, u) = \lambda\eta e^{-(\lambda t+\eta u)}\ \left(1+\rho\{1-2e^{-\lambda t}-2e^{-\eta u}+4e^{-(\lambda t+\eta u)}\}\right) \qquad (14.38)$$

and the marginal densities are the constituent exponentials regardless of the value of ρ.

A second model that is an obvious choice is the bivariate Normal. The density function for this model is well known, so it is not restated here. As is also well known, the marginal densities are Normal.

One final model that we wish to consider here is the one stated by Hunter [75] in a queuing context but also consistent with reliability interpretations:

$$f_{T,U}(t, u) = \frac{\lambda\eta}{1-\rho} I_0\left(\frac{2\sqrt{\rho}}{1-\rho}\sqrt{\lambda\eta t u}\right)\ \exp\left\{-\frac{\lambda t+\eta u}{1-\rho}\right\} \qquad (14.39)$$

where $I_n(\cdot)$ is the modified Bessel function of order n. The marginal densities for this model are not obvious.

To summarize the construction to this point, we have defined examples of two classes of models that might be used to portray the dispersion in equipment longevity as defined using two variables. We next examine the general probability concepts commonly associated with reliability analysis and use some of the suggested example forms to illustrate the concepts discussed. Subsequently, we construct reliability and maintenance models using the probability concepts and some of the example model forms.

14.4.2.3 Probability Analysis

Consider a device for which longevity is defined in terms of two variables, say age and usage. Assuming the distribution function on longevity has been constructed and is bivariate, there are some subtle and sometimes difficult questions and concepts that arise in the application of the bivariate model to reliability. First, we interpret the cumulative failure probability $F_{T,U}(t, u)$ as the probability that failure occurs by time t and usage u. That is,

$$F_{T,U}(t, u) = \Pr[T \leq t, U \leq u] \tag{14.40}$$

One may say that this probability corresponds to the proportion of the population of devices that have longevity vector values at failure that do not exceed (t, u) in either vector component. We emphasize this definition because of the fact that, for a bivariate distribution, probability is generally computed over rectangles such as $[t_1 \leq T \leq t_2, u_1 \leq U \leq u_2]$. Consequently, for any specific longevity vector, (t, u), the range of age and use values implies that there are four rectangles in the (T, U) plane for which probabilities may be meaningfully calculated. Referring to Figure 14.1, observe that, in addition to the rectangle used in Equation 14.40, there are the rectangles $\Pr[T \leq t, U \geq u]$, $\Pr[T \geq t, U \leq u]$, and $\Pr[T \geq t, U \geq u]$. It is not obvious, but relative to the cumulative probability $F_{T,U}(t, u)$, the probabilities

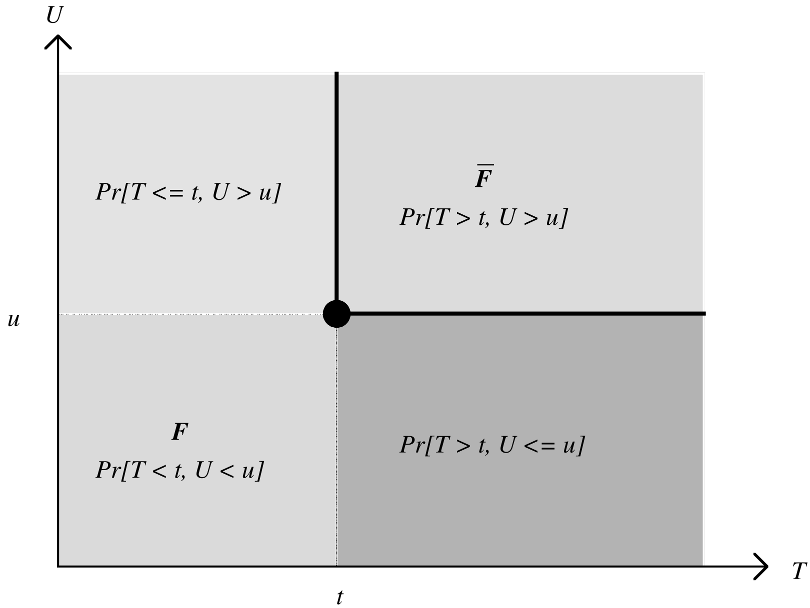


Figure 14.1 Bivaraite probability distributions. (From Yang, S.C. and J.A. Nachlas, "Bivariate Reliability and Maintenance Planning Models." *IEEE Transactions on Reliability*, Vol. 50, no. 1, pp. 26–35, IEEE, 2001. With permission.)

$$\Pr[T \leq t, U > u] = \int_0^t \int_u^\infty f_{T,U}(s, v)dvds \qquad (14.41)$$

and

$$\Pr[T > t, U \leq u] = \int_t^\infty \int_0^u f_{T,U}(s, v)dvds \qquad (14.42)$$

are survival probabilities. They correspond to the proportions of the population that do not have longevity vectors inferior to (t, u), either because their failure ages exceed t, or their failure usages exceed u. We do not have informative names for the probabilities represented by Equations 14.41 and 14.42, but have considered names such as marginal survival probabilities.

A further point that is rather subtle is the fact that the reliability at (t, u) does not include the probabilities represented by Equations 14.41 and 14.42. The reliability at longevity vector value (t, u) corresponds to the proportion of the population for which failure age exceeds t and failure usage exceeds u. Therefore, the reliability function corresponding to $F_{T,U}(t, u)$ is

$$\overline{F}_{T,U}(t, u) = \Pr[T \geq t, U \geq u] = \int_t^\infty \int_u^\infty f_{T,U}(s, v) dv ds \quad (14.43)$$

Because it does not include the probabilities represented by Expressions 14.41 and 14.42, we call this the reliability rather than the survivor function. We also note that the cumulative failure probability and the reliability no longer sum to one.

The apparent paradox in the definitions of $F_{T,U}(t,u)$ and $\overline{F}_{T,U}(t, u)$ arises from distinctions in point of observation. When considering the distribution, all positive valued longevity vectors can potentially occur, and across a population of devices, all do occur. Relative to the distribution, the cumulative probability at (t, u) does not include devices for which either T exceeds t or U exceeds u. On the other hand, all copies of a device population that have achieved a longevity of (t, u) will have longevity vectors at failure that lie within the rectangle $[t \leq T < \infty, u \leq U < \infty]$, so at (t, u) the rectangles corresponding to the marginal survival probabilities are not accessible.

The computation of bivariate probabilities is reasonably clear. For any rectangle, say $[t_1 \leq T \leq t_2, u_1 \leq U \leq u_2]$, in the plane, the probability of observing a failure at a point included in the rectangle is

$$\Pr[t_1 \leq T \leq t_2, u_1 \leq U \leq u_2] = F_{T,U}(t_2, u_2) - F_{T,U}(t_2, u_1) - F_{T,U}(t_1, u_2) + F_{T,U}(t_1, u_1) \quad (14.44)$$

A useful special case of this expression applies to the reliability function, which may be represented by

$$\overline{F}_{T,U}(t,u) = \Pr[t \le T < \infty, u \le U < \infty]$$

$$= 1 - F_U(u) - F_T(t) + F_{T,U}(t,u) \qquad (14.45)$$

Observe that this expression may also be used to compute cumulative probabilities in cases in which the reliability function is easier than the distribution function to analyze.

Very often, the first question that follows the definition of a probability model for device failure is that of the identity and behavior of the associated hazard function. For a bivariate failure distribution, a return to first principles yields

$$z_{T,U}(t,u) = \lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{\Pr[t \le T \le t+\Delta t, u \le U \le u+\Delta u \mid T>t, U>u]}{\Delta u \Delta t}$$

$$= \lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{\Pr[t \le T \le t+\Delta t, u \le U \le u+\Delta u]}{\Delta u \Delta t \Pr[T>t, U>u]}$$

$$= \frac{1}{\overline{F}_{T,U}(t,u)} [\lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{F_{T,U}(t+\Delta t, u+\Delta u)}{\Delta u \Delta t}$$

$$- \lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{F_{T,U}(t+\Delta t, u)}{\Delta u \Delta t} - \lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{F_{T,U}(t, u+\Delta t)}{\Delta u \Delta t}$$

$$+ \lim_{\substack{\Delta t \to 0 \\ \Delta u \to 0}} \frac{F_{T,U}(t,u)}{\Delta u \Delta t}]$$

$$= \frac{f_{T,U}(t,u)}{\overline{F}_{T,U}(t,u)} \qquad (14.46)$$

which is a very appealing result.

Naturally, the next question is whether or not the hazard function is increasing. Barlow and Proschan [11] define MIFR (multivariate increasing failure rate), and the application of that definition to the bivariate life distributions is that a distribution is MIFR if and only if

$$\frac{\bar{F}_{T,U}(t+s,\, u+v)}{\bar{F}_{T,U}(t,\, u)} \tag{14.47}$$

is nonincreasing in (t, u). The same statement applies to the marginal distributions. That is, it must also be the case that

$$\frac{\bar{F}_T(t+s)}{\bar{F}_T(t)} \text{ and } \frac{\bar{F}_U(u+v)}{\bar{F}_U(u)} \tag{14.48}$$

are nonincreasing in t and u, respectively. The application of these conditions is far from direct. Most models require numerical analysis to characterize hazard function behavior, and that behavior may be rather complicated. For example, for some choices of its parameters, the bivariate exponential distribution of Equation 14.39 displays a hazard that is increasing in usage and decreasing in time.

It should be noted that there is an important and subtle difference between the construction of univariate reliability models on the basis of an assumed hazard form and the corresponding model definition for a bivariate longevity distribution. As is well known, an assumed univariate hazard function, $z_T(t)$, directly implies the life distribution function by the relation

$$F_T(t) = 1 - e^{-\int_0^t z_T(x)dx}$$

which follows as the solution of the differential equation:

$$f_T(t) = \frac{dF_T(t)}{dt} = z_T(t)\big(1 - F_T(t)\big)$$

The corresponding bivariate (partial) differential equation is

$$f_{T,U}(t,\, u) = \frac{\partial^2 F_{T,U}(t,\, u)}{\partial t \partial u} = z_{T,U}(t,\, u)\bar{F}_{T,U}(t,\, u) \tag{14.49}$$

The solution for this equation has not yet been found, so one may not build a bivariate reliability model from the hazard as is done in the univariate case.

A further question is how one computes the mean and other descriptive measures for a bivariate longevity distribution. The answer is that, as with univariate distributions, one begins by constructing the moment generating function (or Laplace Transform) and then obtains moments as successive derivatives of the moment-generating function. The moment-generating function for the bivariate failure distribution is

$$\psi_{T,U}(\theta_1, \theta_2) = E[e^{\theta_1 t + \theta_2 u}] \tag{14.50}$$

and its construction is not always simple.

Also of critical importance to bivariate failure modeling is the question how convolutions are constructed and how bivariate renewal functions are defined and interpreted. Fortunately, the convolution theorem has been shown to extend directly to the bivariate case. On the other hand, the definition and interpretation of the associated counting process and the bivariate renewal function is less obvious and may depend upon the application. This is a topic that is treated below and for which considerable further study is needed.

14.4.2.4 Failure and Renewal Models

Consider next the renewal model based on the operation of a sequence of identical copies of a device to failure with instantaneous replacement. For a bivariate longevity distribution, the sequence of device lives forms a bivariate renewal process.

Hunter [75] defines a bivariate renewal process using the vector (X_n, Y_n), where it is assumed that the (T_n, U_n) are independently and identically distributed with the common joint distribution $F_{T,U}(t, u)$, and

$$(X_n, Y_n) = \left(\sum_{i=1}^{n} T_i, \sum_{i=1}^{n} U_i \right) \tag{14.51}$$

Of course, by convention, $(X_0, Y_0) = (0, 0)$. The renewal vector is illustrated in Figure 14.2 below. Referring to the figure, we note that the number of renewals at any coordinate point in the plane, say (t, u), corresponds to the largest value

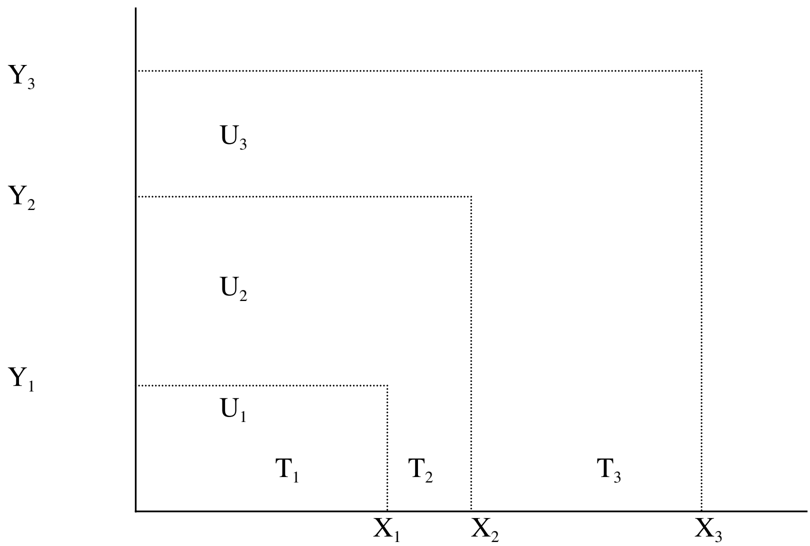


Figure 14.2 A bivariate renewal process. (From Yang, S.C. and J.A. Nachlas, "Bivariate Reliability and Maintenance Planning Models." *IEEE Transactions on Reliability*, Vol. 50, no. 1, pp. 26–35, IEEE, 2001. With permission.)

of n for which the *nth* renewal occurs on or before time t and usage u. Therefore, it follows that the number of renewals by (t, u) is given by

$$N_{T,U}(t,u) = \sup\{n : n \geq 0,\ X_n \leq t,\ Y_n \leq u\} \tag{14.52}$$

Under this definition, $N_{T,U}(t,u)$ is a bivariate renewal counting process for which the distribution is obtained using the usual "time-frequency duality" relation. Hunter confirms that this leads to

$$P[N_{T,U}(t,u) = n] = F_{T,U}^{(n)}(t,u) - F_{T,U}^{(n+1)}(t,u) \tag{14.53}$$

Observe that this construction implies that

$$P[N_{T,U}(t,u) = 0] = 1 - F_{T,U}(t,u) \tag{14.54}$$

which is intuitively appealing. In addition, using Expression 14.53 to obtain the renewal function yields

$$M_F(t, u) = E[N_{T,U}(t, u)] = \sum_{n=1}^{\infty} F^{(n)}(t, u) \qquad (14.55)$$

which corresponds to the univariate form. As in the case of the univariate function, the recursive statement of Expression 14.55 is the *key integral renewal equation*:

$$M_F(t, u) = F_{T,U}(t, u) + \int_0^t \int_0^u M_F(t-x, u-y) dF_{T,U}(x, y) \quad (14.56)$$

and this function is the basis for analysis of the renewal process. As a final point, note that assuming $F_{T,U}(t, u)$ is absolutely continuous implies that the renewal density exists and is

$$m_F(t, u) = \frac{\partial^2}{\partial t \partial u} M_F(t, u) = \sum_{n=1}^{\infty} f_{T,U}^{(n)}(t, u)$$

$$= f_{T,U}(t, u) + \int_0^t \int_0^u m_F(t-x, u-y) f_{T,U}(x, y) dxdy \quad (14.57)$$

The Laplace (or Laplace-Stieltjes) transform serves as the usual method of analysis for the renewal models. For the bivariate case, the Laplace transform of the density associated with the distribution function, $F_{T,U}(t, u)$ is

$$f_{T,U}^*(s, v) = E[e^{-st-vu}] = \int_0^{\infty} \int_0^{\infty} e^{-st-vu} f_{T,U}(t, u) dudt \quad (14.58)$$

and

$$F_{T,U}^*(s, v) = \frac{1}{sv} f_{T,U}^*(s, v) \qquad (14.59)$$

Using these forms in the analysis of the key renewal equation leads to

$$M_F^*(s, v) = \frac{F_{T,U}^*(s, v)}{1 - f_{T,U}^*(s, v)} \quad \text{and} \quad F_{T,U}^*(s, v) = \frac{M_F^*(s, v)}{1 + svM_F^*(s, v)} \quad (14.60)$$

as well as

$$m_F^*(s, v) = \frac{f_{T,U}^*(s, v)}{1 - f_{T,U}^*(s, v)}$$

and

$$f_{T,U}^*(s, v) = \frac{m_F^*(s, v)}{1 + m_F^*(s, v)} \tag{14.61}$$

which correspond to the univariate forms.

Application of these results to specific bivariate distributions is never simple. In the case of the stochastic functions, the algebraic complexity of the models implies that the Laplace transforms cannot be obtained in closed form, but must be computed numerically. Even the numerical construction of the transforms is taxing. The procedure is to select a grid size over the (s, v) plane and to perform the integration of Equation 14.58 numerically. The results may then be stored in an array for further use or fit (approximately) using a second-order regression model.

For the bivariate distributions selected to represent correlation between the variables, expressions for the Laplace transforms can be obtained. For the bivariate exponential distribution shown in Equations 14.37 and 14.38, the renewal function has

$$M_F^*(s, v) = \frac{\lambda\eta\left[\left(v + 2\eta\right)\left(s + 2\lambda\right) + sv\rho\right]}{sv\left[\left(v + 2\eta\right)\left(s + 2\lambda\right)\left(s\left(v + \eta\right) + v\lambda\right) - sv\eta\lambda\rho\right]} \tag{14.62}$$

Similarly, the bivariate exponential distribution of Equation 14.39 has

$$M_F^*(s, v) = \left[sv\left(\frac{s}{\lambda} + \frac{v}{\eta} + \frac{sv}{\lambda\eta}(1 - \rho)\right)\right]^{-1} \tag{14.63}$$

and the Laplace Transform of the renewal function for the bivariate Normal distribution is

$$M_F^*(s,v)=\frac{\exp\left[-s\mu_t-v\mu_u+\frac{1}{2}\begin{pmatrix}s^2\sigma_t^2+2\rho sv\sigma_t\sigma_u\\ +v^2\sigma_u^2\end{pmatrix}\right]}{sv\left[1-\exp\left[-s\mu_t-v\mu_u+\frac{1}{2}\begin{pmatrix}s^2\sigma_t^2+2\rho sv\sigma_t\sigma_u\\ +v^2\sigma_u^2\end{pmatrix}\right]\right]} \tag{14.64}$$

Unfortunately, the inverse transforms for these expressions cannot be constructed in closed form. For Equation 14.63, working with the transform of the associated renewal density permits its inversion [75] to

$$m_f(t,u)=\frac{\lambda\eta}{1-\rho}I_0\left(\frac{2\sqrt{\lambda\eta tu}}{1-\rho}\right)\exp\left\{-\frac{\lambda t+\eta u}{1-\rho}\right\} \tag{14.65}$$

It is rare that the transform is invertible in closed form. For most of the bivariate models, closed form transform inversions are not available.

Next, consider the cases in which repair is no longer instantaneous. Assume instead that repair effort extends over a bivariate interval that is random. Assume further that the distribution function on the magnitude of the repair effort is denoted by $G_r(t, u)$ and is of the same family as the failure distribution. There is no justification for the assumption that the failure and repair distributions are of the same family. The reason for using this assumption here is the fact that it sometimes makes the analysis easier. In addition, we observe that for cases in which it is appropriate, the repair distribution can easily be collapsed into a univariate form.

To construct renewal models for the failure with noninstantaneous repair cases, we obtain the convolution on the operating and repair intervals and then the renewal function based on the convolution. As availability is the quantity of interest for these cases, the renewal function is used to obtain the availability function. That is, considering a longevity cycle to be the sum of an operating interval and a repair interval, the cycle has distribution $H_r(t, u)$, which is obtained as the inverse transform of

$$H_r^*(s, v) = F_{T,U}^*(s, v)G_r^*(s, v) \tag{14.66}$$

Then using the same reasoning as for univariate models, we observe that a device is available at coordinate point (t, u) if it experiences no failures prior to that point, or else it is renewed at some earlier coordinate point and experiences no further failures before (t, u). That is,

$$\mathbf{A}(t, u) = \bar{F}_{T,U}(t, u) + \int_0^t \int_0^u \bar{F}_{T,U}(t-x, u-y)dM_H(x, y) \tag{14.67}$$

where $M_H(t, u)$ is obtained using Equation 14.59. Also comparable to the univariate case is the fact that the availability function is ultimately obtained as the inverse transform of

$$\mathbf{A}^*(s, v) = \bar{F}_{T,U}^*(s, v)\left(1 + m_H^*(s, v)\right) = \frac{\bar{F}_{T,U}^*(s, v)}{1 - h^*(s, v)} \tag{14.68}$$

As indicated previously, inversion of the transform is quite a challenge.

Taking the bivariate exponential density of Equation 14.38 as the failure density and assuming the repair time density comes from the same family means that the repair time density is

$$g_{T,U}(t, u) = \lambda_r \eta_r e^{-(\lambda_r t + \eta_r u)} \left(1 + \rho_r \left\{ \begin{matrix} 1 - 2e^{-\lambda_r t} - 2e^{-\eta_r u} \\ + 4e^{-(\lambda_r t + \eta_r u)} \end{matrix} \right\} \right) \tag{14.69}$$

and

$$h_r^*(s, v) = \frac{\begin{pmatrix} \lambda\lambda_r \eta\eta_r \left((v + 2\eta_r)(s + 2\lambda_r) + sv\rho_r\right) \\ \left((v + 2\eta)(s + 2\lambda) + sv\rho\right) \end{pmatrix}}{\begin{pmatrix} \left((v + \eta_r)(s + \lambda_r)(v + 2\eta_r)(s + 2\lambda_r)\right) \\ \left((v + \eta)(s + \lambda)(v + 2\eta)(s + 2\lambda)\right) \end{pmatrix}} \tag{14.70}$$

so that

$$\mathbf{A}^*(s, v) = \frac{a_1}{a_2 - a_3} \tag{14.71}$$

with

$$a_1 = \big((s+2\lambda)(\nu+2\eta)+\lambda\eta\rho\big)\big((s+\lambda)(\nu+\eta)(s+2\lambda)(\nu+2\eta)\big)$$

$$a_2 = \begin{bmatrix} (s+\lambda_r)(v+\eta_r)(s+2\lambda_r)(v+2\eta_r) \\ (s+\lambda)(v+\eta)(s+2\lambda)(v+2\eta) \end{bmatrix}$$

and

$$a_3 = \lambda\lambda_r\eta\eta_r\big((s+2\lambda_r)(\nu+2\eta_r)s\nu\rho_r\big)\big((s+2\lambda)(\nu+2\eta)+s\nu\rho\big)$$

The corresponding construction for Hunter's bivariate exponential distribution of Equation 14.39 yields

$$h_r^*(s, v) = \left(\frac{\lambda\eta}{(s+\lambda)(\nu+\eta)-s\nu\rho}\right)\left(\frac{\lambda_r\eta_r}{(s+\lambda_r)(\nu+\eta_r)-s\nu\rho_r}\right) \tag{14.72}$$

Then, using the probability identity of Equation 14.45 simplifies slightly the construction of

$$\mathbf{A}^*(s, v) = \frac{\dfrac{1}{sv} - \dfrac{s+\lambda-1}{s(s+\lambda)} - \dfrac{v+\eta-1}{v(v+\eta)} + \dfrac{\lambda\eta}{(s+\lambda)(v+\eta)-sv\rho}}{1-\left(\dfrac{\lambda\eta}{(s+\lambda)(v+\eta)-sv\rho}\right)\left(\dfrac{\lambda_1\eta_1}{(s+\lambda_r)(v+\eta_r)-sv\rho_r}\right)} \tag{14.73}$$

Finally, for the bivariate Normal models, we again use Equation 14.45 with the result that

$$\mathbf{A}^*(s, v) = \frac{1 - vf_T^*(s) - sf_U^*(v) + sf_{T,U}^*(s, v)}{sv\big(1-h^*(s, v)\big)} \tag{14.74}$$

where the joint distribution has

$$f^*_{T,U}(s, v) = \exp\left[-s\mu_t - v\mu_u + \tfrac{1}{2}\left(s^2\sigma_t^2 + 2\rho sv\sigma_t\sigma_u + v^2\sigma_u^2\right)\right] \quad (14.75)$$

the marginal distributions have

$$f^*_T(s) = \exp\left[-\mu_t s + \tfrac{1}{2}\left(s^2\sigma_t^2\right)\right]$$

and

$$f^*_U(v) = \exp\left[-\mu_u v + \tfrac{1}{2}\left(v^2\sigma_u^2\right)\right] \quad (14.76)$$

and

$$h^*_r(s, v) = \exp\left[-s\left(\mu_t + \mu_{t_r}\right) - v\left(\mu_u + \mu_{u_r}\right)\right.$$

$$\left. + \tfrac{1}{2}\left(s^2\left(\sigma_t^2 + \sigma_{t_r}^2\right) + 2sv\rho\left(\sigma_t\sigma_u + \sigma_{t_r}\sigma_{u_r}\right) + v^2\left(\sigma_u^2 + \sigma_{u_r}^2\right)\right)\right] \quad (14.77)$$

Each of the availability function expressions obtained is far too intricate to allow for closed form inversion. Numerical evaluation is difficult but is the only practical approach.

To close this discussion, note that Yang and Nachlas [72] provide the construction of a model of age replacement for the bivariate renewal models. Comparable models for other preventive maintenance cases have not yet been defined, but they can be constructed. Thus, the extension of the traditional univariate maintenance planning models for application to systems for which longevity is bivariate is appropriate and manageable.

14.5 ACKNOWLEDGMENT

Material in Section 14.4.2 through Section 14.4.2.4 is based on Yang, S.C. and J.A. Nachlas, "Bivariate Reliability and Maintenance Planning Models." *IEEE Transactions on Reliability*, Vol. 50, no. 1, pp. 26–35, IEEE, 2001, with permission.

14.6 EXERCISES

1. For a given life distribution and set of repair costs, which type of warranty is more expensive for a manufacturer? Prove your response using expected costs.
2. Suppose a full replacement warranty period is to be set so that no more than 4% of a population of product units are expected to fail during the warranty interval. If the devices have a Weibull life distribution with θ = 200 weeks and β = 2.25, what warranty period (in operating weeks) should be used?
3. Compute the optimal full replacement warranty period for the product of Problem 2 for the case in which $\gamma = 0.1$, $u_l = 40.0$, and $u_2 = 4.0$.
4. Compute the optimal pro rata warranty period for the product of Problem 2 for the case in which $\gamma = 0.75$, $u_l = 40.0$, and $u_2 = 4.0$.

Appendix A

Numerical Approximations

There are many situations in which the analysis of a reliability problem involves numerical calculations. For some of those calculations, precise numerical computation is not possible, but numerical approximations are available. Several useful numerical approximations are provided here. Most of these are based on descriptions provided by Abramowitz and Stegun [76]. If desired, alternate or additional algorithms may be found there.

A.1 NORMAL DISTRIBUTION FUNCTION

For the general Normal distribution with mean μ and variance σ^2, one often denotes the distribution by $N(\mu, \sigma^2)$ and the Standard Normal Distribution is denoted by N(0, 1). In general,

$$\begin{aligned} F_X(x) = N(x, \mu, \sigma^2) &= \int_{-\infty}^{x} f_X(u)du \\ &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(u-\mu)^2/2\sigma^2} du \end{aligned} \tag{A.1.1}$$

and the standard Normal transformation is

$$z = \frac{x - \mu}{\sigma} \tag{A.1.2}$$

It is convention that the Standard Normal Distribution Function is denoted by Φ(z) and that the density function is denoted by ϕ(z). Under the standard transformation,

$$F_X(x) = \Phi\left(z = \frac{x - \mu}{\sigma}\right) \tag{A.1.3}$$

and Abramowitz and Stegun [76] provide the approximation

$$\Phi(z) \approx 1 - \frac{1}{2}\left(1 + d_1x + d_2x^2 + d_3x^3 + d_4x^4 + d_5x^5 + d_6x^6\right)^{-16} + \varepsilon(z) \tag{A.1.4}$$

where
d_1 = 0.04986(73)
d_2 = 0.02114(10)
d_3 = 0.00327(76)
d_4 = 0.00003(80)
d_5 = 0.00004(89)
d_6 = 0.00000(54)

and the calculation yields a result with $|\varepsilon(z)| < 1.5 \times 10^{-7}$. This algorithm was used to generate the entries in Table A.1.

There is also an algorithm for computing quantiles in the tail of the distribution. Specifically, for γ in the range $0 \le \gamma \le 0.05$, Abramowitz and Stegun [76] indicate that

$$z_{1-\gamma} \approx t - \frac{c_0 + c_1t + c_2t^2}{1 + e_1t + e_2t^2 + e_3t^3} + \varepsilon(\gamma) \tag{A.1.5}$$

where
c_0 = 2.515517
c_1 = 0.802853
c_2 = 0.010328
e_1 = 1.432788
e_2 = 0.189269
e_3 = 0.001308

and

$$t = \sqrt{\ln\left(\frac{1}{\gamma^2}\right)} \tag{A.1.6}$$

This approximation yields results for which $|\varepsilon(\gamma)| < 4.5 \times 10^{-4}$.

Table A.1 Standard Normal Cumulative Probabilities

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.86214
1.1	0.8643	0.8665	0.8686	0.8708	0.8728	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9648	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9712	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

A.2 GAMMA FUNCTION

When working with the Gamma distribution, the Weibull distribution, and some other probability functions, one sometimes needs to evaluate a gamma function. In general, the gamma function is defined as the definite integral

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \tag{A.2.1}$$

In the cases in which x is an integer, we know that

$$\Gamma(x) = (x-1)! \tag{A.2.2}$$

For cases in which x is not an integer, Abramowitz and Stegun [76] provide the approximation:

$$\Gamma(x+1) = "x!" \approx 1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5 + \varepsilon(x) \tag{A.2.3}$$

for values of x such that $0 \le x \le 1$. In this expression, we use
$a_1 = -0.5748646$
$a_2 = 0.9512363$
$a_3 = -0.6998588$
$a_4 = 0.4245549$
$a_5 = -0.1010678$

and the resulting error is bounded by $|\varepsilon(x)| < 5 \times 10^{-5}$.

As an example, observe that we compute $\Gamma(5.64)$ as

$$\begin{aligned}\Gamma(5.64) &= (4.64)(3.64)(2.64)(1.64)\,\Gamma(1.64)\\ &= (73.1252)(0.8986) = 65.7122\end{aligned}$$

A.3 PSI (DIGAMMA) FUNCTION

The Psi (or digamma) function is defined as the derivative of the logarithm of the gamma function. It is therefore the ratio of the derivative of the gamma function and the gamma function:

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \left. \frac{d}{dx}\Gamma(x) \middle/ \Gamma(x) \right. \tag{A.3.1}$$

Tables of the Psi function are available [76]. On the other hand, approximate numerical evaluation of the Psi function is reasonably straightforward except for the fact that precise values may require that a large number of terms be included in the calculation.

The general series expansion for the Psi function is

$$\psi(x+1) = -\gamma + \sum_{n=2}^{\infty} (-1)^n \zeta(n) x^{n-1}$$

$$|x| < 1 \tag{A.3.2}$$

where

$$\zeta(n) = \sum_{k=1}^{\infty} k^{-n} \tag{A.3.3}$$

is the Riemann zeta function, and γ is Euler's constant, for which the value is $\gamma = 0.57721(56649)$. For values of the argument of the function that exceed 1, we use the recursion:

$$\psi(x+1) = \psi(x) + \frac{1}{x} \tag{A.3.4}$$

Combining the above expressions implies that a numerical approximation for any Psi function may be computed by truncating the infinite sums. As stated above, it appears that the truncation must be made after rather many terms to obtain good precision, but modern computing power makes this manageable.

Appendix B

Numerical Evaluation of Weibull Renewal Functions

The Weibull distribution provides a very useful model of life lengths. The utility of the distribution results from its tractability when treating simple probabilities for life lengths. When exploring renewal behavior, it is less manageable. In fact, there are no closed form expressions for the convolutions or the renewal function for the Weibull distribution. Lomnicki [49] shows that the Weibull renewal function has an equivalent representation as the sum of terms of a MacLauren series, and that the same is true for convolutions of a Weibull distribution. Specifically, suppose we represent the Weibull distribution function as

$$F_T(t) = 1 - e^{-\left(t/\theta\right)^{\beta}} \tag{B.1}$$

Then, define the "Poissonian function":

$$P_k\left(\left(t/\theta\right)^{\beta}\right) = \frac{\left(\left(t/\theta\right)^{\beta}\right)^k}{k!} e^{-\left(t/\theta\right)^{\beta}} \tag{B.2}$$

with the corresponding remainder accumulation:

$$D_k\left(\left(t/\theta\right)^{\beta}\right) = \sum_{j=k}^{\infty} P_j\left(\left(t/\theta\right)^{\beta}\right) \tag{B.3}$$

Then, the Weibull renewal function is represented as the infinite sum:

$$M_{F_T}(t) = \sum_{k=1}^{\infty} c(k) D_k\left(\left(t/\theta\right)^{\beta}\right) \tag{B.4}$$

for which the coefficients, $c(k)$, are computed as indicated by Lomnicki and explained below. Observe that the corresponding representation of the k-fold convolution of the Weibull distribution is

$$F_T^{(k)}(t) = \sum_{j=k}^{\infty} \phi_j(k) D_j\left(\left(t/\theta\right)^{\beta}\right) \tag{B.5}$$

in which the coefficients $\phi_j(k)$ are also defined by Lomnicki and shown below. Before showing the calculation of the coefficients, we note that the renewal density and the convolution of the probability density may also be obtained by taking the derivatives of the functions in Expressions B.4 and B.5, respectively. This is shown here.

Taking the derivatives of the renewal and convolution distribution functions, we find that both depend primarily on the derivative of the function D_k, which comprises the key challenge. That is,

$$\begin{aligned} m_{F_T}(t) &= \frac{d}{dt} M_{F_T}(t) = \frac{d}{dt} \sum_{k=1}^{\infty} c(k) D_k\left(\left(t/\theta\right)^{\beta}\right) \\ &= \sum_{k=1}^{\infty} c(k) \frac{d}{dt} D_k\left(\left(t/\theta\right)^{\beta}\right) \end{aligned} \tag{B.6}$$

$$\begin{aligned} f_T^{(k)}(t) &= \frac{d}{dt} F_T^{(k)}(t) = \frac{d}{dt} \sum_{j=k}^{\infty} \phi_j(k) D_j\left(\left(t/\theta\right)^{\beta}\right) \\ &= \sum_{j=k}^{\infty} \phi_j(k) \frac{d}{dt} D_j\left(\left(t/\theta\right)^{\beta}\right) \end{aligned} \tag{B.7}$$

Given the definition of the function D_k, we have

$$\frac{d}{dt}D_k\left(\left(\frac{t}{\theta}\right)^{\beta}\right)=\frac{d}{dt}\sum_{j=k}^{\infty}P_j\left(\left(\frac{t}{\theta}\right)^{\beta}\right)=\sum_{j=k}^{\infty}\frac{d}{dt}P_j\left(\left(\frac{t}{\theta}\right)^{\beta}\right)$$

$$=\sum_{j=k}^{\infty}\frac{d}{dt}\left(\frac{\left(\left(\frac{t}{\theta}\right)^{\beta}\right)^{j}}{j!}e^{-\left(\frac{t}{\theta}\right)^{\beta}}\right)=\sum_{j=k}^{\infty}\frac{d}{dt}\left(\frac{\left(\frac{t}{\theta}\right)^{j\beta}}{j!}e^{-\left(\frac{t}{\theta}\right)^{\beta}}\right)$$

$$=\sum_{j=k}^{\infty}\left(\frac{\left(\frac{j\beta}{t}\right)\left(\frac{t}{\theta}\right)^{j\beta}}{j!}-\frac{\left(\frac{\beta}{t}\right)\left(\frac{t}{\theta}\right)^{\beta}\left(\frac{t}{\theta}\right)^{j\beta}}{j!}\right)e^{-\left(\frac{t}{\theta}\right)^{\beta}}$$

$$=\frac{\beta}{t}\sum_{j=k}^{\infty}\left(\frac{\left(\frac{t}{\theta}\right)^{j\beta}}{(j-1)!}-\frac{\left(\frac{t}{\theta}\right)^{(j+1)\beta}}{j!}\right)e^{-\left(\frac{t}{\theta}\right)^{\beta}}$$

$$=\frac{\beta}{t}\frac{\left(\frac{t}{\theta}\right)^{k\beta}}{(k-1)!}e^{-\left(\frac{t}{\theta}\right)^{\beta}} \tag{B.8}$$

Substituting this result into Expressions B.6 and B.7, we obtain

$$m_{F_T}(t)=\sum_{k=1}^{\infty}c(k)\frac{\beta}{t}\frac{\left(\frac{t}{\theta}\right)^{k\beta}}{(k-1)!}e^{-\left(\frac{t}{\theta}\right)^{\beta}} \tag{B.9}$$

and

$$f_T^{(k)}(t)=\sum_{j=k}^{\infty}\phi_j(k)\frac{\beta}{t}\frac{\left(\frac{t}{\theta}\right)^{j\beta}}{(j-1)!}e^{-\left(\frac{t}{\theta}\right)^{\beta}} \tag{B.10}$$

Thus, all of the interesting probability measures can be computed using the coefficients defined by Lomnicki.

To obtain those coefficients, we start by computing a ratio of gamma functions:

$$\gamma(k) = \Gamma(k\beta+1) \big/ \Gamma(k+1) \tag{B.11}$$

Then, we apply the recursive definition:

$$\begin{aligned} b_0(k) &= \gamma(k) \\ b_{j+1}(k) &= \sum_{i=j}^{k-1} b_j(i)\gamma(k-i) \end{aligned} \tag{B.12}$$

for which $j \geq 0$ and $k \geq j+1$. We use the resulting values to obtain

$$a_j(k) = \sum_{i=j}^{k} (-1)^{j+i} \binom{s}{i} \frac{b_j(i)}{\gamma(i)} \tag{B.13}$$

and then,

$$\begin{aligned} \phi_j(j) &= a_j(j) \\ \phi_j(k) &= \sum_{i=j}^{k} a_i(k) - \sum_{i=j}^{k-1} a_i(k-1) \end{aligned} \tag{B.14}$$

where again, $k \geq j+1$. These are the coefficients we use to compute the convolutions on the distribution and density functions. Finally, we obtain the coefficients for the renewal function and the renewal density:

$$c(k) = \sum_{j=1}^{k} \phi_j(k) \tag{B.15}$$

To obtain a sense of the computations involved and the likely accuracy of the finite sum of terms, consider a set of example calculations. Suppose $\beta = 1.5$, and since the scale is arbitrary, take $\theta = 1$. Using these values, the first 16 of the gamma-function-based quantities are given in Table B.1.

Table B.1 Values of $\gamma(k)$ for $\beta = 1.5$

k	γ	k	γ
0	1.0000	8	11880.0000
1	1.3293	9	63636.2377
2	3.0000	10	360360.0000
3	8.7238	11	2.1453×10^6
4	30.0000	12	13.3661×10^6
5	116.9534	13	86.8191×10^6
6	504.0000	14	586.0512×10^6
7	2360.9966	15	4099.8766×10^6

Then, using the quantities, $\gamma(k)$, we use Expression B.12 to obtain the $b_j(k)$, shown in Table B.2.

Next, the analysis of Expression B.13 yields the results in Table B.3 and Table B.4.

Finally, we obtain the values for the coefficients $c(k)$, which are given in Table B.5.

Table B.2 Values of the Quantities $b_j(k)$ for $\beta = 1.5$

$k\backslash j$	0	1	2	3	4	5
0	1.0000					
1	1.3293	1.3293				
2	3.0000	4.7672	1.7671			
3	8.7238	16.6998	10.3251	2.3491		
4	30.0000	62.1938	48.0981	19.0271	3.1228	
5	116.9534	249.0566	214.2441	110.3305	32.3409	4.1513
6	504.0000	1071.0464	961.8336	572.1867	224.2416	52.3605
7	2360.9966	4926.1199	4442.0375	2857.3636	1328.0853	422.3589
8	11880.0000	24121.6916	21294.4744	14200.6517	7323.0231	2814.0220
9	63636.2377	125198.1354	106397.8593	71453.2323	39160.3981	17010.7346
10	360360.0000	685940.2457	555031.7309	367598.2560	207719.7852	97696.0360
11	2.1453×10^6	3.9521×10^6	3.0421×10^6	1.9452×10^6	1.1080×10^6	547237.1877
12	13.3661×10^6	23.8614×10^6	17.2026×10^6	10.6275×10^6	5.9966×10^6	3.0392×10^6
13	86.8191×10^6	150.4915×10^6	102.0784×10^6	60.0923×10^6	33.1335×10^6	16.9205×10^6
14	586.0512×10^6	988.5588×10^6	631.0661×10^6	352.1645×10^6	187.7382×10^6	95.1867×10^6
15	4099.8766×10^6	6745.3736×10^6	4058.6121×10^6	2140.5447×10^6	1094.3582×10^6	544.2424×10^6

$k\backslash j$	6	7	8	9	10
6	5.5185				
7	82.0588	7.3359			
8	754.7551	125.6394	9.7519		
9	5589.1903	1297.6446	189.0252	12.9636	
10	36796.0159	10575.6191	2165.9264	280.5346	17.2330
11	226339.1912	75373.4950	19267.6607	3531.4026	411.8167
12	1.3390×10^6	495049.8362	147871.4150	34052.6332	5649.1317
13	7.7593×10^6	3.0903×10^6	1.0338×10^6	280080.8014	58697.9832
14	44.6066×10^6	18.7034×10^6	6.8004×10^6	2.0780×10^6	515220.0330
15	2.567160×10^6	111.2747×10^6	4.29887×10^6	14.3811×10^6	4.0449×10^6

$k\backslash j$	11	12	13	14	15
11	22.9086				
12	599.1438	30.4533			
13	8895.4067	865.1917	40.4828		
14	99086.5916	13822.3045	1241.4942	53.8154	
15	924648.5878	164320.0939	21235.7913	1771.8168	71.5390

Table B.3 Values of the Quantities $a_j(k)$ for $\beta = 1.5$

$k\backslash j$	0	1	2	3	4	5	6	7
0	1.0000							
1	0.0000	1.0000						
2	0.0000	0.4110	0.5890					
3	0.0000	0.1471	0.5836	0.2693				
4	0.0000	0.0497	0.4033	0.4429	0.1041			
5	0.0000	0.0163	0.2393	0.4650	0.2439	0.0355		
6	0.0000	0.0052	0.1306	0.3970	0.3472	0.1091	0.0109	
7	0.0000	0.0017	0.0677	0.3005	0.3881	0.1971	0.0419	0.0031
8	0.0000	0.0005	0.0339	0.2102	0.3752	0.2731	0.0921	0.0143
9	0.0000	1.60×10^{-4}	0.0165	0.1392	0.3288	0.3211	0.1525	0.0371
10	0.0000	4.96×10^{-5}	0.0079	0.0885	0.2686	0.3375	0.2114	0.0715
11	0.0000	1.52×10^{-5}	0.0037	0.0546	0.2081	0.3267	0.2588	0.1142
12	0.0000	4.66×10^{-6}	0.0017	0.0328	0.1547	0.2971	0.2892	0.1597
13	0.0000	1.42×10^{-6}	0.0008	0.0194	0.1113	0.2570	0.3011	0.2021
14	0.0000	4.32×10^{-7}	0.0004	0.0112	0.0779	0.2137	0.2963	0.2367
15	0.0000	1.31×10^{-7}	0.0002	0.0064	0.0534	0.1719	0.2785	0.2605

$k \backslash j$	8	9	10	11	12	13	14	15
8	0.0008							
9	0.0044	0.0002						
10	0.0132	0.0013	4.78×10^{-5}					
11	0.0292	0.0043	0.0003	1.07×10^{-5}				
12	0.0528	0.0106	0.0013	8.33×10^{-5}	2.28×10^{-6}			
13	0.0830	0.0215	0.0035	0.0004	1.97×10^{-5}	4.66×10^{-7}		
14	0.1173	0.0375	0.0079	0.0011	9.14×10^{-5}	4.41×10^{-6}	9.18×10^{-8}	
15	0.1525	0.0586	0.0151	0.0026	0.0003	2.23×10^{-5}	9.45×10^{-7}	1.74×10^{-8}

Table B.4 Values of the Quantities $\phi_j(k)$ for $\beta = 1.5$

$k \backslash j$	1	2	3	4	5	6	7	8
0								
1	1.0000							
2	0.0000	0.5890						
3	0.0000	0.2638	0.2693					
4	0.0000	0.0974	0.2776	0.1041				
5	0.0000	0.0334	0.1974	0.1753	0.0355			
6	0.0000	0.0111	0.1197	0.1878	0.0845	0.0109		
7	0.0000	0.0036	0.0665	0.1630	0.1220	0.0340	0.0031	
8	0.0000	0.0011	0.0350	0.1252	0.1382	0.0622	0.0120	0.0008
9	0.0000	0.0003	0.0177	0.0887	0.1351	0.0870	0.0266	0.0038
10	0.0000	0.0001	0.0087	0.0594	0.1196	0.1033	0.0444	0.0099
11	0.0000	3.43×10^{-5}	0.0042	0.0382	0.0987	0.1095	0.0620	0.0193
12	0.0000	1.06×10^{-5}	0.0020	0.0237	0.0771	0.1068	0.0764	0.0310
13	0.0000	3.24×10^{-6}	0.0009	0.0144	0.0578	0.0978	0.0860	0.0436
14	0.0000	9.89×10^{-7}	0.0004	0.0086	0.0419	0.0853	0.0901	0.0555
15	0.0000	3.01×10^{-7}	0.0002	0.0050	0.0296	0.0714	0.0891	0.0653

$k\backslash j$	9	10	11	12	13	14	15
9	0.0002						
10	0.0011	1.07×10^{-5}					
11	0.0033	0.0002	1.07×10^{-5}				
12	0.0074	0.0010	7.49×10^{-5}	2.28×10^{-6}			
13	0.0134	0.0025	0.0003	1.78×10^{-5}	4.66×10^{-7}		
14	0.0212	0.0051	0.0008	7.57×10^{-5}	4.04×10^{-6}	9.18×10^{-8}	
15	0.0300	0.0090	0.0018	0.0002	1.88×10^{-5}	8.71×10^{-7}	1.74×10^{-8}

Table B.5 Values of $c(k)$ for $\beta = 1.5$

k	$c(k)$	k	$c(k)$
1	1.0000	9	0.3595
2	0.5890	10	0.3466
3	0.5331	11	0.3354
4	0.4792	12	0.3256
5	0.4417	13	0.3168
6	0.4140	14	0.3089
7	0.3923	15	0.3017
8	0.3744		

Appendix C

Laplace Transform for the Key Renewal Theorem

In Chapter 9, the *Key Renewal Theorem* is presented, and its Laplace Transform is stated to be

$$M_{F_T}^*(s) = F_T^*(s) + L\left(\int_0^t M_{F_T}(t-u) f_T(u)du\right) = F_T^*(s) + M_{F_T}^*(s) f_T^*(s) \quad \text{(C.1)}$$

To see that this is correct, start with the general definition of the transform:

$$L\big(F_T(t)\big) = \int_0^\infty e^{-st} F_T(t)dt = F_T^*(s) \qquad \text{(C.2)}$$

Then, we apply the definition to the renewal function, and the transform of the integral term is

$$L\left(\int_0^t M_{F_T}(t-u) f_T(u)du\right) = \int_0^\infty e^{-st} \int_0^t M_{F_T}(t-u) f_T(u)dudt$$

$$= \int_0^\infty \int_0^t e^{-s(t-u)} \sum_{n=1}^\infty F_T^{(n)}(t-u) e^{-su} f_T(u)dudt$$

$$= \int_0^\infty \int_0^t \sum_{n=1}^\infty e^{-s(t-u)} F_T^{(n)}(t-u) e^{-su} f_T(u)dudt$$

$$= \int_0^\infty \int_u^\infty \sum_{n=1}^{\infty} e^{-s(t-u)} F_T^{(n)}(t-u) e^{-su} f_T(u)\,dt\,du$$

$$= \sum_{n=1}^{\infty} \int_0^\infty \left(\int_u^\infty e^{-s(t-u)} F_T^{(n)}(t-u)\,dt \right) e^{-su} f_T(u)\,du$$

$$= \sum_{n=1}^{\infty} \int_0^\infty F_T^{(n)*}(s) e^{-su} f_T(u)\,du$$

$$= \sum_{n=1}^{\infty} F_T^{(n)*}(s) \int_0^\infty e^{-su} f_T(u)\,du$$

$$= \sum_{n=1}^{\infty} F_T^{(n)*}(s) f_T^*(s) = f_T^*(s) \sum_{n=1}^{\infty} F_T^{(n)*}(s)$$

$$= M_{F_T}^*(s) f_T^*(s) \tag{C.3}$$

Bibliography

1. Schlager, N. (1994). *When Technology Fails*. Detroit, Gale Research, Inc.
2. NASA. (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident*. NASA Report No. S677. Washington, DC, U.S. Government Printing Office.
3. NASA. (2003). *Columbia Accident Investigation Board Report*. Washington, DC, U.S. Government Printing Office.
4. Power System Outage Task Force, North America Electric Reliability Council. (2004). *Final Report on the August 14, 2003 Blackout in the US and Canada: Causes and Recommendations*. Princeton, NJ, NERC.
5. Blanchard, B. S. (1992). *Logistics Engineering and Management*. Englewood Cliffs, NJ, Prentice Hall.
6. Thompkins, J. A. (1989). *Winning Manufacturing*. New York, McGraw Hill.
7. Shetty, Y. K., Buehler, V. M. (eds.) (1988). *Competing Through Productivity and Quality.* Cambridge, MA, Productivity Press.
8. Nachlas, J. A., Cassady, C. R. (1999). Preventive Maintenance Study: A Key Component in Engineering Education to Enhance Industrial Productivity and Competitiveness. *Eur. J. Eng. Educ.* 24, 299–309.
9. Dougherty, Jr., E. M., Fragola, J. R. (1988). *Human Reliability Analysis: A Systems Engineering Approach with Nuclear Power Plant Applications*. New York, John Wiley & Sons.

10. Musa, J. D., Iannino, A., Okumoto, K. (1987). *Software Reliability: Measurement, Prediction, Application*. New York, McGraw Hill.
11. Barlow, R. E., Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. New York, Holt, Reinhart, and Winston.
12. Natvig, B. (1982). Two Suggestions of How to Define a Multistate Coherent System. *Adv. Appl. Probab*. 14, 434–455.
13. Barlow, R. E., Wu, A. S. (1978). Coherent Systems with Multistate Components. *Math. Operations Res*. 4, 275–281.
14. Majety, S. R., Dawande, M., Rajgopal, J. (1999). Optimal Reliability Allocation with Discrete Cost-Reliability Data for Components. *Operations Res*. 47, 899–906.
15. Rice, W. F., Cassady, C. R., Wise, T. R. (1999). Simplifying the Solution of Redundancy Allocation Problems. *Proc. Annual Reliability and Maintainability Symposium*, pp. 190–194.
16. Lloyd, D. K., Lipow, M. (1962). *Reliability: Management, Methods and Mathematics*. Englewood Cliffs, NJ, Prentice Hall.
17. Wong, K. L., Lindstrom, D. L. (1988). Off the Bathtub onto the Rollercoaster Curve. *Proc. Annual Reliability and Maintainability Symposium*, pp. 356–363.
18. Barlow, R. E., Marshall, A. W., Proschan, F. (1963). Properties of Probability Distriburion with Monotone Hazard Rate. *Ann. Math.l Stat*. 34, 375–389.
19. Weibull, W. (1939). A Statistical Theory of the Strength of Materials. *Proc. of the Swedish Royal Institute of Engineering Research,* p. 153.
20. Weibull, W. (1951). A Statistical Distribution Function of Wide Applicability. *J. Appl. Mech*. 18, 293–297.
21. Gumbel, E. J. (1958). *Statistics of Extremes*. New York, Columbia University Press.
22. Fisher, R. A., Tippet, L. H. C. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proc. of the Cambridge Philosophical Soc.* 24, 180–190.

23. Birnbaum, Z. W. (1958). A Statistical Model for Life Length of Materials. *J. Am. Stat. Assoc.* 53, 151–160.

24. Birnbaum, Z. W., Saunders, S. C. (1969). A New Family of Life Distributions. *J. Appl. Probab*. 6, 319–327.

25. Jordan, C. W. (1967). *Life Contingencies*. Chicago, Society of Actuaries.

26. Hjorth, U. (1980). A Reliability Distribution with Increasing, Decreasing and Bathtub Shaped Failure Rates. *Technometrics* 22, 99–107.

27. Kapur, K. C., Lamberson, L. R. (1977). *Reliability in Engineering Design*. New York, John Wiley & Sons.

28. Gertsbakh, I. H., Kordonskiy, K. H. (1969). *Models of Failure*. Berlin, Springer Verlag.

29. Karlin, S. (1964). Total Positivity, Absorbtion Probabilities and Applications. *Trans. Am. Math. Soc*. 3, 33–107.

30. Nachlas, J. A. (1988). A Stochastic Model of Component Failure. *Proc. of the Sixth International Conf. on Reliability and Maintainability*, pp. 487–491.

31. Arrhenius, S. Z. (1989). *Phys. Chem*. 4, 120

32. Krausz, A. S., Eyring, H. B. (1975). *Deformation Kinetics*. New York, John Wiley & Sons.

33. Stevenson, J. L., Nachlas, J. A. (1990). Microelectronics Reliability Predictions Derived From Component Defect Densities. *Proc. Annual Reliability and Maintainability Symposium*, pp. 366–371.

34. Kumar, D., Klefsjo, B. (1994). Proportional Hazards Model: A Review. *Reliability Eng. Syst. Saf.* 44, 177–188.

35. Lemoine, A. J., Wenocur, M. L. (1985). On Failure Modeling. *Nav. Res. Logistics Q*. 32, 497–508.

36. Cox, D. R., Oakes, D. (1984). *Analysis of Survival Data*. New York, Chapman and Hall.

37. Nachlas, J. A. (1986). A General Model for Age Acceleration During Thermal Cycling. *Qual. Reliability Eng. Int*. 2, 3–6.

38. Jensen, F. (1985) Activation Energies and the Arrhenius Equation, *Quality and Reliability International* 1, 13–17.

39. Vassilou, P., Mettas, A. (2004). Understanding Accelerated Life Testing Analysis. *Topics in Reliability—Tutorial Notes. Annual Reliability and Maintainability Symposium*.

40. Military Standard 781. *Reliability Design Qualification and Production Acceptance Tests*. Washington, DC, Naval Electronic Systems Command.

41. Barlow, R. E., Campo, R. (1975). Total Time on Test Processes and Applications to Failure Data Analysis. In *Reliability and Fault Tree Analysis*. (R. E. Harlow, J. Fussel, N. D. Singpurwalla, eds.). Philadelphia, SIAM, pp. 451–481.

42. Klefsjo, B. (1982). On Aging Properties and Total Time on Test Transforms. *Scand. J. Stat*. 9, 37–41.

43. Nachlas, J. A., Kumar, A. (1993). Reliability Estimation Using Doubly Censored Field Data. *IEEE Trans. Reliability* 42, 268–279.

44. Cox, D. R. (1962). *Renewal Theory*. London. Methuen.

45. Nelson, W. (1980). Accelerated Life Testing — Step-Stress Models and Data Analysis. *IEEE Trans. Reliability* 29, 103–108.

46. Nelson, W. (1990). *Accelerated Testing: Statistical Models, Test Plans and Data Analysis*. New York, John Wiley & Sons.

47. Meeker, W. Q., Escobar, L. A. (1998). *Statistical Methods for Reliability Data*. New York, John Wiley & Sons.

48. Feller, W. (1966). *An Introduction to Probability Theory and Its Applications. II*. New York, John Wiley & Sons.

49. Lomnicki, Z. A. (1966). A Note on the Weibull Renewal Process. *Biometrika* 53, 375–381.

50. Barlow, R. E., Hunter, L. C. (1960). Optimum Preventive Maintenance Policies. *Operations Res*. 8, 90–100.

51. Brown, M., Proschan, F. (1983). Imperfect Repair. *J. Appl. Probab*. 20, 851–859.

52. Block, H. W., Borges, W. S., Savits, T. H. (1985). Age Dependent Minimal Repair. *J. Appl. Probab*. 22, 370–385.

53. Kijima, M. (1989). Some Results for Repairable Systems with General Repair. *J. Appl Probab*. 26, 89–102.

54. Kijima, M., Morimura, H., Suzuki, Y. (1988). Periodical Replacement Problem without Assuming Minimal Repair. *Eur. J. Operations Res*. 37, 194–203.

55. Wang, H., Pham, H. (1996). Optimal Maintenance Policies for Several Imperfect Repair Models. *Int. J. Syst. Sci*. 27, 543–549.

56. Wang, H., Pham, H. (1996). A Quasi Renewal Process and Its Applications in Imperfect Maintenance. *Int. J. Syst. Sci*. 27, 1055–1062.

57. Cassady, C. R., Nachlas, J. A. (1994). The Frequency Distribution of Availability. *Proc. Annual Reliability and Maintainability Symposium*, pp. 278–282.

58. Iyer, S. (1992). Availability Results for Imperfect Repair. *Sankhaya, Indian J. Stat*. 54.B, 249–256.

59. Birolini, A. (1985). *On the Use of Stochastic Processes in Modeling Reliability Problems*. Lecture Notes in Ecomonics and Mathematical Systems, p. 252, Berlin, Springer Verlag.

60. Ross, S. M. (1988). *A First Course in Probability*. New York, Macmillan.

61. Murdock, W. P., Nachlas, J. A. (2003). Time Dependent Availability under Age Replacement with Distinct Service Time Distributions. *Proc. of the 17th International Conference on Production Research*.

62. Degbotse, A., Nachlas, J. A. (2003). Use of Nested Renewals to Model Availability Under Opportunistic Maintenance Policies. *Proc. of the Annual Reliability and Maintainability Symposium*. pp. 344–350.

63. Nakagawa, T. (1986). Periodic and Sequential Preventive Maintenance Policies. *J. Appl. Probab*. 23, 536–542.

64. Block, H. W., Borges, W. S., Savits, T. H. (1986). Preventive Maintenance Policies. In *Reliability and Quality Control* (A. P. Basu, ed.), Amsterdam, Elsevier Science Publishers, pp. 101–106.

65. Makis, V., Jardine, A. K. S. (1993). A Note On Optimal Replacement Policy Under General Repair. *Eur. J. Operations Res*. 69, 75–82.

66. van Noortwijk, J. M., Cooke, R., Kok, M. (1995). A Bayesian Failure Model Based On Isotropic Deterioration. *Eur. J. Operations Res*. 82, 270–282.

67. Grall, A., Dieulle, L., Berenguer, C., Roussignol, M. (2002). Continuous Time Predictive Maintenance Scheduling for a Deteriorating System. *IEEE Trans. Reliability* 51, 141–150.

68. Duane, J. T. (1964). Learning Curve Approach to Reliability Monitoring. *IEEE Trans. Aerosp.* 2, 563–566.

69. Crow, L. H. (1984). Methods for Assessing Reliability Growth Potential. *Proc. of the Annual Reliability and Maintainability Symposium*, pp. 484–489.

70. U.S. Department of Defense. (1981). *Military Handbook 189: Reliability Growth Management*.

71. Singpurwalla, N. D., Wilson, S. P. (1998). Failure Models Indexed by Two Scales. *Adv. Appl. Probab*. 30, 1–58–1072.

72. Yang, S. C., Nachlas, J. A. (2001). Bivariate Reliability and Maintenance Planning Models. *IEEE Trans. Reliability* 50, 26–35.

73. Eliashberg, J., Singpurwalla, N. D., Wilson, S. P. (1997). Calculating the Reserve for a Time and Usage Indexed Warranty. *Manage. Sci*. 43, 966–975.

74. Baggs, G. E., Nagagaja, H. N. (1996). Reliability Properties of Order Statistics From Bivariate Exponential Distributions. *Commun. Stat*. 12, 611–631.

75. Hunter, J. (1974). Renewal Theory in Two Dimensions: Basic Results, *Adv. Appl. Probab*. 6, 376–391.

76. Abramowitz, M., Stegun, I. A. (1965). *Handbook of Mathematical Functions*. New York, Dover Publications.

Index

B

C

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W