Contributions to Statistics

Elias Ould Saïd
Idir Ouassou
Mustapha Rachdi *Editors*

Functional Statistics and Applications

Selected Papers from MICPS-2013

Springer

Contributions to Statistics

More information about this series at http://www.springer.com/series/2912

Elias Ould Saïd • Idir Ouassou • Mustapha Rachdi
Editors

Functional Statistics and Applications

Selected Papers from MICPS-2013

Springer

Editors
Elias Ould Saïd
LMPA
Université du Littoral Côte d'Opale
Calais, France

Idir Ouassou
National School of Applied Sciences
Cadi Ayyad University
Marrakesh, Morocco

Mustapha Rachdi
Faculty of Science of Man and Society
Grenoble-Alpes University
Grenoble, France

ISSN 1431-1968
Contributions to Statistics
ISBN 978-3-319-22475-6 ISBN 978-3-319-22476-3 (eBook)
DOI 10.1007/978-3-319-22476-3

Library of Congress Control Number: 2015957212

Mathematics Subject Classification (2010): 60H30, 60H07, 62G08, 62G20, 92B15, 60H40

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This volume, dedicated to recent advances in statistical methodology and applications, has two parts. The first is devoted to functional statistics, which involves taking a mathematical function (usually a real vector function of real scalar or vector variables, like time and/or space) as an entity to globally estimate and test in a suitable functional space. The approach is often nonparametric and involves the approximation of the studied function by the element of a functional kernel, then the calculation of some characteristics (temporal and/or spatial) of this approximant, considered as estimators of the homologous characteristics of the studied function and, finally, to study the asymptotic distribution and convergence properties of these estimators. This first part includes the following themes: the M-estimation of the regression function for quasi-associated processes, by Said Attaoui, Ali Laksaci, and Elias Ould Said; the kernel estimation of extreme conditional quantiles, by Stéphane Girard and Sana Louhichi; the estimation of a linear regression operator for functional covariates, by Amina Naceri, Ali Laksaci, and Mustapha Rachdi; and the estimation of a loss function for spherically symmetric distribution, by Idir Ouassou.

For the second part, three application areas are preferred. First area comes from the tradition of signal analysis inaugurated by J. Fourier (see Fig. 1) in Grenoble two centuries ago in his famous paper on the propagation of heat in solid bodies, published in the *Nouveau Bulletin des Sciences de la Société Philomatique de Paris* (1, 112–116, 1808). After his return to France from the Egyptian campaign in 1802, Napoleon appointed J. Fourier as "préfet" of Grenoble and he recreates in 1810 as rector, the University of Grenoble, who had already started in 1339 under the auspices of Dauphin Humbert II, with the approval of Pope Benedict XII (14th founded medieval European University). J. Fourier will host in this university as Professor JF Champollion, the father of Egyptology. This first area concerns the following contributions: the approximation of strictly stationary Banach-valued random sequence by Fourier integral, by Tawfik Benchikh; the proposal of a new tool for biological signal processing, the Dynalets, a natural generalization of the Fourier and wavelet transforms, by Jacques Demongeot, Ali Hamie, Olivier Hansen,

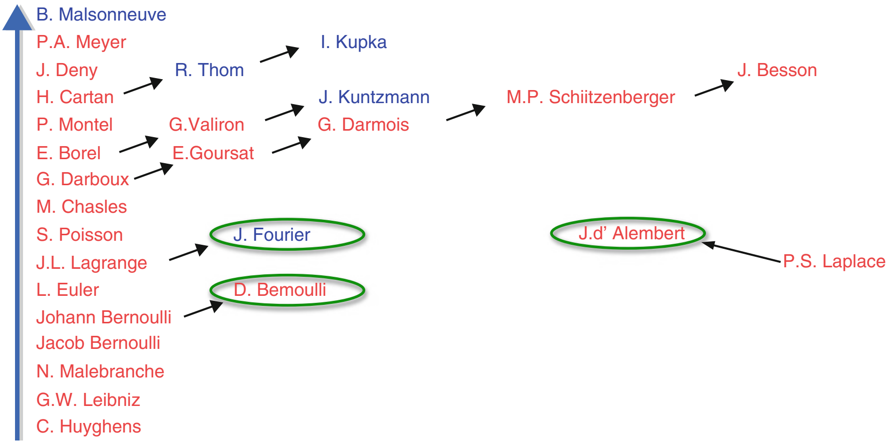

Fig. 1 Genealogy (from AMS Mathematics Genealogy Project) of B. Maisonneuve, author of the regenerative systems theory (PhD supervisor is located on the line below his student name, with, in *blue*, Grenoble scientists)

and Mustapha Rachdi; and Estimation of the block shrink wavelet density in φ-mixing framework, by M. Badaoui and N. Rhomari.
The second application area revisits the classic approach initiated by D. Bernoulli, mathematician and physician, prominent member of the large family of Swiss mathematicians, the Bernoulli family, and J. d'Alembert (son of the famous Madame du Tencin from Grenoble), by using copula theory, for estimating some characteristics of the temporal evolution of subpopulations involved in the epidemic spread of a vector-borne disease in the copula approach in epidemiologic modeling, by J. Demongeot, M. Ghassani, H. Hazgui, and M. Rachdi.

The third application domain concerns estimation problems in queuing theory in the contribution to the impact of nonparametric density estimation on the approximation of the G/G/1 Queue, by A. Bareche and D. Aissani, and stochastic analysis of an M/G/1 Retrial Queue, by M. Boualem, M. Cherfaoui, N. Djellab, and D. Aissani.

Grenoble, France
September 2014

Jacques Demongeot

Acknowledgements

This book could not have been published without the assistance of many people and organizations.

We would like to express our gratitude to Prof. Y. Ouknine (resident member of the Hassan II Academy of Science and Technology), as well as to Profs. M. Eddahbi and K. Essebaiy for their dedication in the bi-annual organization of the Marrakesh International Conference of Probability and Statistics, which has been a phenomenal achievement.

The Hassan II Academy of Science and Technology has funded much of the MCPS'13. We would like to thank the academy for their support.

We would like to acknowledge all the contributors for providing their researches that helped prepare this book.

We are also indebted to anonymous proofreaders and reviewers.

We are grateful to the Springer staff and especially to Dr. Veronika Rosteck for her attention and care in the preparation of this manuscript, as well as for her patience.

Grenoble, France — Mustapha Rachdi
Calais, France — Elias Ould Saïd
Marrakesh, Morocco — Idir Ouassou

Contents

Contributors

Djamil Aïssani
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Exact Sciences, University of Bejaia, Bejaia, Algeria

Said Attaoui
Département de Mathématiques, Université Djillali Liabès, Sidi Bel Abbès, Algeria

Mohammed Badaoui
LaMSD et URAC 04, Département de Mathématiques et Informatique, Faculté des Sciences, Université Mohamed 1er, Oujda, Morocco

Aïcha Bareche
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Technology, University of Bejaia, Bejaia, Algeria

Tawfik Benchikh
Département de Mathématiques, Univ. Djillali Liabès, Sidi Bel Abbès, Algeria

Mohamed Boualem
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Technology, University of Bejaia, Bejaia, Algeria

Mouloud Cherfaoui
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Exact Sciences, University of Bejaia, Bejaia, Algeria

Jacques Demongeot
Univ. Grenoble-Alpes, AGIM FRE 3405 CNRS, Faculty of Medicine, University J. Fourier of Grenoble, La Tronche, France

Natalia Djellab
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Exact Sciences, University of Bejaia, Bejaia, Algeria

Mohamad Ghassani
Univ. Grenoble-Alpes, AGIM FRE 3405 CNRS, Faculty of Medicine, University J. Fourier of Grenoble, La Tronche, France

Stephane Girard
Mistis, Inria Grenoble Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, France

Ali Hamie
Univ. Grenoble-Alpes, AGIM FRE 3405 CNRS, Faculty of Medicine, University J. Fourier of Grenoble, La Tronche, France

Olivier Hansen
Univ. Grenoble-Alpes, AGIM FRE 3405 CNRS, Faculty of Medicine, University J. Fourier of Grenoble, La Tronche, France

Hana Hazgui
Univ. Grenoble-Alpes, AGIM FRE 3405 CNRS, Faculty of Medicine, University J. Fourier of Grenoble, La Tronche, France

Ali Laksaci
Département de Mathématiques, Univ. Djillali Liabès, Sidi Bel Abbès, Algeria

Sana Louhichi
Laboratoire Jean Kuntzmann, Univ. Grenoble-Alpes, Saint-Martin-d'Hères, France

Amina Naceri
Département de Mathématiques, Univ. Djillali Liabès, Sidi Bel Abbès, Algeria

Idir Ouassou
National School of Applied Sciences, Cadi Ayyad University, Marrakesh, Morroco

Elias Ould Saïd
Université Lille Nord de France, Lille, France

ULCO, LMPA, Calais, France

Mustapha Rachdi
Univ. Grenoble-Alpes, Laboratoire AGIM FRE 3405 CNRS, UPMF, UFR SHS, Grenoble, France

Nouredine Rhomari
LaMSD et URAC 04, Département de Mathématiques et Informatique, Faculté des Sciences, Université Mohamed 1er, Oujda, Morocco

Part I
Functional Statistics

Asymptotic Results for an M-Estimator of the Regression Function for Quasi-Associated Processes

Said Attaoui, Ali Laksaci, and Elias Ould Saïd

Abstract In this paper, we study a family of robust nonparametric estimators for the regression function based on the kernel method. It is assumed that the observations form a stationary quasi-associated sequence. Under general conditions we establish the almost-complete convergence with rate of the estimator as well as its asymptotic normality.

1 Introduction

Let $(Z_i)_{i=1,\ldots,n} := (X_i, Y_i)_{i=1,\ldots,n}$ be an $\mathbb{R}^d \times \mathbb{R}$-valued measurable and strictly stationary process, defined on the probability space $(\Omega, \mathscr{A}, \mathbb{P})$. Our purpose is to study the co-variation between X_i and Y_i via the robust estimation of the regression function. This nonparametric model, denoted by θ_x, is implicitly defined as a zero with respect to (w.r.t.) t, in the equation:

$$\Psi(x,t) := \mathbb{E}\left[\psi(Y_1, t) \mid X_1 = x\right] = 0, \tag{1}$$

S. Attaoui (✉)
Département de Mathématiques, Université Djillali Liabès, BP 89, 22000 Sidi Bel Abbès, Algerie

Département de Mathématiques, Université des Sciences et de la Technologie, Mohamed Boudiaf, BP 1505, 31000 El Mnaouer-Oran, Algerie
e-mail: s_attaoui@yahoo.fr

A. Laksaci
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, BP 89, 22000 Sidi Bel Abbès, Algerie
e-mail: alilak@yahoo.fr

E. Ould Said
Université Lille Nord de France, F-59000 Lille, France

ULCO, LMPA, F-62228 Calais, France
e-mail: ouldsaid@lmpa.univ-littoral.fr

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_1

where ψ is a real-valued Borel function satisfying some regularity conditions to be stated below. We suppose that, for all $x \in \mathbb{R}^d$, θ_x exists and is unique (cf. for instance, [2]). A natural estimator of θ_x denoted by $\widehat{\theta_x}$ is a zero w.r.t. t of the equation:

$$\hat{\Psi}(x,t) = 0 \tag{2}$$

with

$$\hat{\Psi}(x,t) := \frac{\sum_{i=1}^{n} K(h^{-1}(x - X_i))\psi(Y_i, t)}{\sum_{i=1}^{n} K(h^{-1}(x - X_i))},$$

where K is a kernel function and $h := h_n$ is a sequence of positive real numbers which goes to zero as n goes to infinity.

It is well known that robust regression is an important analysis tool in statistics. It is used to circumvent some of the limitations of the classical regression, namely when data are heteroscedastic or contain outliers. Due to this interesting feature, robust regression has been widely considered in time series analysis. Key references on this subject are [2, 3, 7, 13–15, 22] for previous results and [4, 18] for recent advances and references. However, in all these works the dependence structure is modeled on the mixing hypothesis. In this paper, we focus on a more general correlation type, that is the quasi-association condition. This kind of dependence structure was introduced by Bulinski and Suquet [6] for real-valued random fields as a generalization of positively associated variables introduced by Esary et al. [12] and negatively associated random variables considered by Jong-Dev and Proschan [16]. Both types of association have great importance in various applied fields (see the book by Barlow and Proschan [1] for a deeper discussion on this topic).

Nonparametric estimation involving (positively and negatively) associated random variables has been extensively studied. We quote, for instance, [20, 21, 23–25] and the reference therein. We refer the reader to [8] or [10] for some other weak dependence structures and their applications.

The goal of this paper is to study a family of nonparametric robust estimators of the regression function, based on the kernel method. These estimates are constructed by combining the ideas of robustness with those of smoothed regression which allows us to obtain reliable estimation when outlier observations are present within the responses. Under general conditions, we establish the uniform almost complete convergence of these estimators and we show their asymptotic normality suitably normalized. As far as we know, only the recent paper by Douge [9] has paid attention to studying the kernel estimation under the quasi-association condition. He studied the asymptotic properties of the kernel estimator of the classical regression which can be viewed as a particular case of the present work.

The paper is organized as follows: the next section is dedicated to fixing notations and hypotheses. We state the uniform almost complete convergence in Sect. 3. The asymptotic normality is given in Sect. 4. The last section is devoted to the proofs.

2 Notations and Hypotheses

We begin by recalling the definition of the quasi-association property.

Definition 1 A sequence $(X_n)_{n\in\mathbb{N}}$ of real random vectors is said to be quasi-associated, if for any disjoint subsets I and J of $\mathbb{N}$ and all bounded Lipschitz functions $f : \mathbb{R}^{|I|d} \to \mathbb{R}$ and $g : \mathbb{R}^{|J|d} \to \mathbb{R}$ we have

$$\left|\mathrm{Cov}(f(X_i,\ i\in I),\ g(X_j,\ j\in J))\right| \leq \mathrm{Lip}(f)\mathrm{Lip}(g)\sum_{i\in I}\sum_{j\in J}\sum_{k=1}^{d}\sum_{l=1}^{d}\left|\mathrm{Cov}(X_i^k, X_j^l)\right| \tag{3}$$

(here and in the sequel $|I|$ denotes cardinality of a finite set I), where X_i^k denotes the kth component of X_i, and

$$\mathrm{Lip}(f) = \sup_{x\neq y}\frac{|f(x)-f(y)|}{\|x-y\|_1},\ \text{ with } \|(x_1,\ldots,x_k)\|_1 = |x_1| + \cdots + |x_k|.$$

Note that there are interesting stochastic models in mathematical statistics, reliability theory, percolation theory, and statistical mechanics described by families of positively or negatively associated random variables (see for instance [6]). It is shown in [5] that positively and negatively associated random variables with finite second moment do satisfy (3). Moreover, the quasi-association property of a Gaussian process $X = \{X_t,\ t\in T\}$ is studied by Shashkin [26].

Throughout this paper, we suppose that the sequence $(X_i, Y_i)_{i=1,\ldots,n}$ is quasi-associated, S is a fixed compact subset of $\mathbb{R}^d$ and f (resp. $f_{i,j}$) the density of X (resp. the joint density of (X_i, X_j)). Furthermore, we set by C or C' some positive generic constants and by

$$\lambda_k := \sup_{s>k}\sum_{|i-j|\geq s}\lambda_{i,j},$$

where

$$\lambda_{i,j} = \sum_{k=1}^{d}\sum_{l=1}^{d}|\mathrm{Cov}(X_i^k, X_j^l)| + \sum_{k=1}^{d}|\mathrm{Cov}(X_i^k, Y_j)| + \sum_{l=1}^{d}|\mathrm{Cov}(Y_i, X_j^l)| + |\mathrm{Cov}(Y_i, Y_j)|.$$

In order to state the asymptotic proprieties of our estimate we consider the following assumptions:

(U1) The density f is of class $\mathscr{C}^1(\mathbb{R}^d)$, such that $\inf_{x\in S} f(x) > C > 0$ and the joint density $f_{i,j}$ satisfies $\sup_{|i-j|\geq 1} \|f_{(X_i,X_j)}\|_\infty < \infty$, where $\|\cdot\|_\infty$ is the supremum norm.

(U2) There exists $\delta_0 > 0$ such that

$$\sup_{x\in S} |\theta_x| \leq \delta_0.$$

(U3) The process $\{(X_i, Y_i),\ i \in \mathbb{N}\}$ is quasi-associated with covariance coefficient $\lambda_k,\ k \in \mathbb{N}$ satisfying

$$\exists a > 0 \ \text{ such that } \ \lambda_k \leq Ce^{-ak}.$$

(U4) $\begin{cases} \text{(U4a) The function } \Psi(.,.) \text{ is of class } \mathscr{C}^1 \text{ on } S\times[-\delta_0,+\delta_0], \text{ such that} \\ \inf_{x\in S,\, t\in[-\delta_0,\delta_0]} \frac{\partial\Psi}{\partial t}(x,t) > C > 0, \\ \text{(U4b) For each fixed } t\in[-\delta_0,+\delta_0], \text{ the function } \Psi(.,t) \text{ is continous} \\ \text{over S.} \end{cases}$

(U5) The function ψ is strictly monotonic w.r.t. the second component, Lipschitz and such that, $\forall t \in [-\delta_0, +\delta_0]$,

$$\mathbb{E}\left(\exp\left(|\psi(Y,t)|\right)\right) \leq C \quad \text{and } \forall i \neq j,\ \mathbb{E}\left(\left|\psi(Y_i,t)\psi(Y_j,t)\right| \Big| X_i, X_j\right) \leq C'.$$

(U6) K is β-Holder with compact support.

(U7) There exist $\gamma \in (0,1)$, $\xi_1, \xi_2 > 0$ and $\gamma + \xi_2 < 1$ such that $\dfrac{C(\log n)^{1/d}}{n^{(1-\gamma-\xi_2)/d}} \leq h \leq \dfrac{C}{(\log n)^{(1+\xi_1)/d}}.$

2.1 Comments on the Hypotheses

Our assumptions are classical for time series analysis. Moreover, the robustness of our model is exploited with Hypothesis (U5), where we keep the same condition given by Collomb and Härdle [7] in the multivariate case. We point out that our robustness condition is verified for the usual functions (Huber, Hample, Tuckey, ...) which gives more flexibility in the practical choice of ψ. Note that the nonparametric regression studied by Douge [9] can be treated as a particular case of our study by taking $\psi(Y,t) = Y - t$. The rest of hypotheses are technical conditions imposed for the concision of proofs. Finally, it should be noted that Condition (U7) is needed to express the of our robust model even if ψ is unbounded. In particular, Condition (U7) yields that $\lim_{n\to\infty} \frac{\log n}{n^{1-\gamma}h^d} = 0$ which implies $\lim_{n\to\infty} \frac{\log n}{nh^d} = 0$.

3 Results

3.1 Consistency

The following result ensures the almost complete consistency of $\hat{\theta}_x$.

Theorem 1 *Under Hypotheses (U1)–(U7), the estimator $\hat{\theta}_x$ exists and is unique. Moreover, we have, as $n \longrightarrow \infty$,*

$$\sup_{x\in S}\left|\hat{\theta}_x-\theta_x\right| = O\left(h+\left(\frac{\log n}{n^{1-\gamma}h^d}\right)^{1/2}\right) \qquad \text{a.co.} \tag{4}$$

3.2 Asymptotic Normality

Now we study the asymptotic normality of $\widehat{\theta_x}$ for a fixed $x \in S$. In this case, Assumption (U7) must be replaced by

(U7'). There exists $\gamma \in (0,1)$ and $\xi_1, \xi_2 > 0$ such that $\frac{1}{n^{2(1/2-\gamma/6-\xi_2)/d}} \leq h \leq \frac{C}{n^{(1+\xi_1)/(d+2)}}$.

Theorem 2 *Assume that (U1)–(U7') hold, then we have for any $x \in S$,*

$$\left(\frac{nh^d}{\sigma^2(x,\theta_x)}\right)^{1/2}\left(\widehat{\theta_x}-\theta_x\right) \xrightarrow{\mathscr{D}} \mathscr{N}(0,1) \quad \text{as} \quad n\to\infty,$$

where

$$\sigma^2(x,\theta_x) = \frac{\mathbb{E}[\psi_x^2(Y,\theta_x)|X=x]}{\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2}\int_{BBr^d} K^2(z)dz,$$

$$\mathscr{A} = \left\{x\in S,\ \mathbb{E}[\psi_x^2(Y,\theta_x)|X=x]\frac{\partial}{\partial t}\Psi(x,\theta_x)\neq 0\right\}$$

and $\xrightarrow{\mathscr{D}}$ denotes the convergence in distribution.

4 Proofs

In the following, we denote, for all $x \in \mathbb{R}^d$ and $i = 1, \ldots, n$ $K_i(x) = K(h^{-1}(x-X_i))$.

Proof of Theorem 1 The proof is based on the fact that ψ is strictly monotonic w.r.t. the second component. However, for sake of simplification, we give only the proof for the increasing case. Under this assumption, we write

$$\sup_{x\in S}|\widehat{\theta_x}-\theta_x| = \sup_{x\in S}|\widehat{\theta_x}-\theta_x|\,\mathbb{1}_{\left\{\sup_{x\in S}|\widehat{\theta_x}|\leq\delta_0\right\}} + \sup_{x\in S}|\widehat{\theta_x}-\theta_x|\,\mathbb{1}_{\left\{\sup_{x\in S}|\widehat{\theta_x}|>\delta_0\right\}}.$$

So, to prove the result, it suffices to prove that

$$\sum_n \mathbb{P}\left(\inf_{x\in S}\widehat{\theta_x} < -\delta_0\right) < \infty, \qquad \sum_n \mathbb{P}\left(\sup_{x\in S}\widehat{\theta_x} > \delta_0\right) < \infty \tag{5}$$

and

$$\sup_{x\in S}\left|\widehat{\theta_x}-\theta_x\right|\,\mathbb{1}_{\left\{\sup_{x\in S}|\widehat{\theta_x}|\leq\delta_0\right\}} = O\left(h+\left(\frac{\log n}{n^{1-\gamma}\,h^d}\right)^{\frac{1}{2}}\right), \qquad \text{a.co.} \tag{6}$$

Since $\hat{\Psi}(x,\cdot)$ is increasing for each $x\in S$, we need to show that

$$\sum_n \mathbb{P}\left(\sup_{x\in S}\hat{\Psi}(x,-\delta_0) > 0\right) < \infty \quad \text{and} \quad \sum_n \mathbb{P}\left(\inf_{x\in S}\hat{\Psi}(x,\delta_0) < 0\right) < \infty.$$

Assumption (U2) implies

$$\sup_{x\in S}\Psi(x,-\delta_0) < 0 \quad \text{and} \quad \inf_{x\in S}\Psi(x,\delta_0) > 0.$$

Provided we can check that

$$\sup_{x\in S}\hat{\Psi}(x,-\delta_0) \longrightarrow \sup_{x\in S}\Psi(x,-\delta_0), \qquad \text{a.co.}$$

and

$$\inf_{x\in S}\hat{\Psi}(x,\delta_0) \longrightarrow \inf_{x\in S}\Psi(x,\delta_0), \qquad \text{a.co.}$$

we obtain

$$\begin{aligned}
&\sum_n \mathbb{P}\left(\sup_{x\in S}\hat{\Psi}(x,-\delta_0) > 0\right) \\
&\quad \leq \sum_n \mathbb{P}\left(\left|\sup_{x\in S}\hat{\Psi}(x,-\delta_0) - \sup_{x\in S}\Psi(x,-\delta_0)\right| \geq \epsilon_1\right) < \infty
\end{aligned}$$

and

$$\sum_n \mathbb{P}\left(\inf_{x\in S}\hat{\Psi}(x,\delta_0)<0\right)\leq \sum_n \mathbb{P}\left(\left|\inf_{x\in S}\hat{\Psi}(x,\delta_0)-\inf_{x\in S}\Psi(x,\delta_0)\right|\geq \epsilon_2\right)<\infty$$

with $\epsilon_1 = -\sup_{x\in S}\Psi(x,-\delta_0)$ and $\epsilon_2 = \inf_{x\in S}\Psi(x,\delta_0)$.

Moreover, under (U4a), we write

$$\left(\widehat{\theta_x}-\theta_x\right)\mathbb{I}_{\left\{|\widehat{\theta_x}-\theta_x|\leq\delta\right\}} = \frac{\Psi(x,\widehat{\theta_x})-\hat{\Psi}(x,\widehat{\theta_x})}{\frac{\partial\Psi}{\partial t}(x,\xi_n)}\mathbb{I}_{\left\{|\widehat{\theta_x}-\theta_x|\leq\delta\right\}},$$

where ξ_n is between $\widehat{\theta_x}$ and θ_x.

Thus, all what is left to show is the convergence rate of

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}|\hat{\Psi}(x,t)-\Psi(x,t)|. \tag{7}$$

The proof of (7) is based on the following decomposition

$$\begin{aligned}
&\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}(x,t)-\Psi(x,t)\right| \\
&\leq \frac{1}{\inf_{x\in S}\left|\hat{\Psi}_D(x)\right|}\left\{\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N(x,t)-\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]\right|\right. \\
&\quad+\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]-H(x,t)\right| \\
&\quad+\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\Psi(x,t)\left(f(x)-\mathbb{E}\left[\hat{\Psi}_D(x)\right]\right)\right| \\
&\quad\left.+\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\Psi(x,t)\left(\mathbb{E}\left[\hat{\Psi}_D(x)\right]-\hat{\Psi}_D(x)\right)\right|\right\},
\end{aligned}$$

where

$$\hat{\Psi}_N(x,t) := \frac{1}{nh^d}\sum_{i=1}^n K\left(h^{-1}(x-X_i)\right)\psi_x(Y_i,t)$$

$$\hat{\Psi}_D(x) := \frac{1}{nh^d}\sum_{i=1}^n K\left(h^{-1}(x-X_i)\right)$$

and $\quad H(x,t) := \Psi(x,t)f(x).$

Finally, the proof of Theorem 1 is based on the following lemmas and corollary.

Lemma 1 *Under Hypotheses (U1), (U4) and (U6), we have, as $n \longrightarrow \infty$*

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]-H(x,t)\right| = O(h).$$

Proof of Lemma 1 By equidistribution of the variables, we get

$$\begin{aligned}\mathbb{E}[\hat{\Psi}_N(x,t)] &= \frac{1}{h^d}\int_{\mathbb{R}^d}\mathbb{E}[\psi(Y,t)|X=u)]K\left(\frac{x-u}{h}\right)f(u)\,du\\ &= \frac{1}{h^d}\int_{\mathbb{R}^d}\Psi(u,t)K\left(\frac{x-u}{h}\right)f(u)\,du\\ &= \int_{\mathbb{R}^d}H(x-hz,t)K(z)\,dz.\end{aligned}$$

Since both f and Ψ are of class $\mathscr{C}^1$, a Taylor expansion of $H(x-hz,t)$ permits to write

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}[\hat{\Psi}_N(x,t)]-H(x,t)\right| = O(h).$$

Lemma 2 *Under Hypotheses (U1), (U3)–(U7), we have*

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N(x,t)-\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]\right| = O\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right) \qquad \text{a.co.}$$

Proof of Lemma 2 Since ψ may not be bounded we employ a truncation method by introducing the following random variable

$$\hat{\Psi}_N^*(x,t) = \frac{1}{nh^d}\sum_{i=1}^{n}K\left(h^{-1}(x-X_i)\right)\psi_x(Y_i,t)\,\mathbb{1}_{|\psi(Y_i,t)|<\mu_n} \text{ with } \mu_n = n^{\gamma/6}.$$

Then, the claimed result is a consequence of the following intermediate results

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}[\hat{\Psi}_N^*(x,t)]-\mathbb{E}[\hat{\Psi}_N(x,t)]\right| = O\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right), \tag{8}$$

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x,t)-\hat{\Psi}_N(x,t)\right| = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right), \tag{9}$$

and

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x,t)-\mathbb{E}[\hat{\Psi}_N^*(x,t)]\right| = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right). \tag{10}$$

We start by proving (10). Since S is compact we write

$$S\subset\bigcup_{j=1}^{d_n}B(x_k,\tau_n),$$

with $d_n = O\left(n^{\beta}\right)$ and $\tau_n = O\left(d_n^{-1}\right)$ where $\beta = \frac{\delta(d+2)}{2}+\frac{1}{2}+\frac{\gamma}{6}$ and $\delta \le (1-\gamma-\xi_2)/d$. Now, for all $x\in S$, let

$$k(x) = \arg\min_{k\in\{1,\ldots d_n\}}\|x-x_k\|,$$

and we consider the following decomposition

$$\begin{aligned}
&\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x,t)-\mathbb{E}\Big[\hat{\Psi}_N^*(x,t)\Big]\right| \\
&\le \underbrace{\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x,t)-\hat{\Psi}_N^*(x_{k(x)},t)\right|}_{T_1} \\
&\quad+\underbrace{\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x_{k(x)},t)-\mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)},t)\right]\right|}_{T_2} \\
&\quad+\underbrace{\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\Big[\hat{\Psi}_N^*(x_{k(x)},t)\Big]-\mathbb{E}\Big[\hat{\Psi}_N^*(x,t)\Big]\right|}_{T_3}.
\end{aligned}$$

- Firstly, for T_2 we use the compactness of $[-\delta_0,\ \delta_0]$ and we write

$$[-\delta_0,\ \delta_0]\subset\bigcup_{j=1}^{z_n}\left(t_j-l_n,t_j+l_n\right) \tag{11}$$

  with $l_n = n^{-1/2}$ and $z_n = O\left(n^{1/2}\right)$. Set

$$\mathscr{G}_n = \left\{t_j-l_n,t_j+l_n,1\le j\le z_n\right\}. \tag{12}$$

By the monotony of $\mathbb{E}[\hat{\Psi}_N^*(x,\cdot)]$ and $\hat{\Psi}_N^*(x,\cdot)$ we obtain, for $1 \le j \le z_n$

$$\mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t_j - l_n)\right] \le \sup_{t\in(t_j-l_n,t_j+l_n)} \mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t)\right] \le \mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t_j + l_n)\right]$$

$$\hat{\Psi}_N^*(x_{k(x)}, t_j - l_n) \le \sup_{t\in(t_j-l_n,t_j+l_n)} \hat{\Psi}_N^*(x_{k(x)}, t) \le \hat{\Psi}_N^*(x_{k(x)}, t_j + l_n). \tag{13}$$

Moreover, by Assumption (U5), we have, for any $t_1, t_2 \in [-\delta_0, \delta_0]$

$$\left|\mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t_1)\right] - \mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t_2)\right]\right| \le C|t_1 - t_2|. \tag{14}$$

So, we deduce from (11)–(14) that

$$\begin{aligned}
&\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x_{k(x)}, t) - \mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)}, t)\right]\right| \\
&\le \max_{1\le k\le d_n}\max_{1\le j\le z_n}\max_{t\in\{t_j-l_n,t_j+l_n\}}\left|\hat{\Psi}_N^*(x_k, t)\right. \\
&\quad \left.- \mathbb{E}\left[\hat{\Psi}_N^*(x_k, t)\right]\right| + 2Cl_n.
\end{aligned}$$

A simple algebraic calculation gives us

$$l_n = o\left(\sqrt{\frac{\log n}{nh^d}}\right). \tag{15}$$

Then, it suffices to prove that for some positive real η sufficiently large

$$\max_{1\le k\le d_n}\max_{1\le j\le z_n}\max_{t\in\{t_j-l_n,t_j+l_n\}}\left|\hat{\Psi}_N^*(x_k, t) - \mathbb{E}\left[\hat{\Psi}_N^*(x_k, t)\right]\right| = O_{\text{a.co.}}\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right). \tag{16}$$

To do that, we use a Bernstein-type inequality for dependent random variables (cf. [17]). Indeed, we put:

$$\hat{\Psi}_N^*(x_k, t)) - \mathbb{E}\left[\hat{\Psi}_N^*(x_k, t))\right] = \sum_{i=1}^{n} \Delta_i,$$

where

$$\Delta_i = \frac{1}{nh^d}\chi(X_i, Y_i),$$

with

$$\chi(u, v) = \psi(v, t)K(h^{-1}(x_k - u))\mathbb{1}_{(|\psi(v,t)|<\mu_n)} \\ -\mathbb{E}\left[\psi(Y_1, t)K(h^{-1}(x_k - X_1))\mathbb{1}_{(|\psi(Y_1,t)|<\mu_n)}\right], \; u \in \mathbb{R}^d, \; v \in \mathbb{R}.$$

Clearly,

$$\|\chi\|_\infty \leq C\mu_n\|K\|_\infty \text{ and}$$
$$\mathrm{Lip}(\chi) \leq (\|K\|_\infty \mathrm{Lip}(\psi) + \mu_n h^{-1}\mathrm{Lip}(K)) \leq C\mu_n h^{-1}\mathrm{Lip}(K).$$

Note that, Kallabis and Newmann's inequality is based on the asymptotic evaluation of $\mathrm{Var}\left(\sum_{i=1}^n \Delta_i\right)$ and $\mathrm{Cov}(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})$, for all $(s_1, \ldots, s_u) \in \mathbb{N}^u$ and $(t_1, \ldots, t_v) \in \mathbb{N}^v$. We start by studying the variance term,

$$\mathrm{Var}\left(\sum_{i=1}^n \Delta_i\right) = n\mathrm{Var}\,(\Delta_1) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \mathrm{Cov}(\Delta_i, \Delta_j). \tag{17}$$

Under (U5), we have

$$\begin{aligned} \mathrm{Var}\,(\Delta_1) &\leq \frac{1}{n^2h^{2d}}\mathbb{E}[|\psi(Y_1, t)K_1(x_k)|^2] \leq C'\frac{1}{n^2h^{2d}}\mathbb{E}[|K_1(x_k)|^2] \\ &\leq C'n^{-2}h^{-d}\int_{\mathbb{R}^d} K^2(u)f(x_k - hu)du \\ &= O\left(n^{-2}h^{-d}\right). \end{aligned} \tag{18}$$

Now, let us evaluate the asymptotic behavior of the sum in the right-hand side of (17). For this we use the technique developed by Masry [19]. Indeed, we need the following decomposition

$$\sum_{i=1}^n \sum_{\substack{j=1 \\ j\neq i}}^n \mathrm{Cov}(\Delta_i, \Delta_j) = \sum_{i=1}^n \sum_{\substack{j=1 \\ 0<|i-j|\leq m_n}}^n \mathrm{Cov}(\Delta_i, \Delta_j) + \sum_{i=1}^n \sum_{\substack{j=1 \\ |i-j|>m_n}}^n \mathrm{Cov}(\Delta_i, \Delta_j),$$

where (m_n) is a sequence of positive integer tending to infinity as n goes to infinity. For $|i-j| \leq m_n$, we use (U1), (U5), and (U6) to write that

$$\begin{aligned} \mathbb{E}[\left|\Delta_i\Delta_j\right|] &\leq C\frac{1}{n^2h^{2d}}\left(\mathbb{E}\left[\left|\psi(Y_i, t)K_i(x_k)\psi(Y_j, z)K_j(x_k)\right|\right]\right. \\ &\quad \left. + \left(\mathbb{E}\left[|\psi(Y_1, t)K_1(x_k)|\right]\right)^2\right) \\ &\leq C\frac{1}{n^2h^{2d}}\left(\mathbb{E}[K_i(x_k)K_j(x_k)] + (\mathbb{E}[K_1(x_k)])^2\right) \end{aligned}$$

$$\leq \frac{1}{n^2}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(u)K(v)f_{(X_i,X_j)}(x_k-hu,x_k-hv)dudv\right.$$
$$\left.+\left(\int_{\mathbb{R}^d} K(u)f(x_k-hu)du\right)^2\right)$$
$$= O\left(n^{-2}\right).$$

Then, we get

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\0<|i-j|\leq m_n}}^{n} \mathrm{Cov}(\Delta_i,\Delta_j) \leq nm_n\left(\mathbb{E}[\Delta_i\Delta_j]\right).$$
$$\leq Cn^{-1}m_n.$$

On the other hand, for $|i-j| > m_n$, we use the quasi-association of the sequence (X_i, Y_i) and (U3), to write

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>m_n}}^{n} \mathrm{Cov}(\Delta_i,\Delta_j) \leq \mu_n^2 n^{-2}h^{-2(d+1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\|i-j|>m_n}}^{n} \lambda_{i,j}$$
$$\leq \mu_n^2 n^{-1}h^{-2(d+1)}\lambda_{m_n}$$
$$\leq \mu_n^2 n^{-1}h^{-2(d+1)}e^{-am_n}.$$

So,

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\i\neq j}}^{n}\mathrm{Cov}(\Delta_i,\Delta_j) \leq C\left(n^{-1}m_n + \mu_n^2 n^{-1}h^{-2(d+1)}e^{-am_n}\right).$$

Take $m_n = \frac{1}{a}\log\left(a\mu_n^2 h^{-2(d+1)}\right)$. Then, by the left part of (U7), we get

$$nh^d\sum_{i=1}^{n}\sum_{\substack{j=1\\i\neq j}}^{n}\mathrm{Cov}(\Delta_i,\Delta_j) \to 0. \tag{19}$$

Finally, by combining (18) and (19), we get:

$$\mathrm{Var}\left(\sum_{i=1}^{n}\Delta_i\right) = O\left(\frac{1}{nh^d}\right).$$

Now, we deal with the covariance term in (17), for all $(s_1, \ldots, s_u) \in \mathbb{N}^u$ and $(t_1, \ldots, t_v) \in \mathbb{N}^v$. To do that, we consider the following cases:

- If $t_1 = s_u$, we obtain

$$
\begin{aligned}
|\mathrm{Cov}(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})| &\leq \left(\frac{C\mu_n \|K\|_\infty}{nh^d}\right)^{u+v} \mathbb{E}\left|K_1^2(x_k)\right| \\
&\leq h^d \left(\frac{C\mu_n}{nh^d}\right)^{u+v}. \qquad (20)
\end{aligned}
$$

- If $t_1 > s_u$, we use the quasi-association condition we get

$$
\begin{aligned}
&|\mathrm{Cov}(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})| \\
&\leq \left(\frac{\mu_n h^{-1}\mathrm{Lip}(K))}{nh^d}\right)^2 \left(\frac{2\mu_n \|K\|_\infty}{nh^d}\right)^{u+v-2} \sum_{i=1}^{u}\sum_{j=1}^{v} \lambda_{s_i,t_j} \\
&\leq h^{-2}\left(\frac{C\mu_n}{nh^d}\right)^{u+v} v\lambda_{t_1-s_u} \\
&\leq h^{-2}\left(\frac{C\mu_n}{nh^d}\right)^{u+v} ve^{-a(t_1-s_u)}. \qquad (21)
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
|\mathrm{Cov}(\Delta_{s_1} \ldots \Delta_{s_u}, \Delta_{t_1} \ldots \Delta_{t_v})| &\leq \left(\frac{2C\mu_n \|K\|_\infty}{nh^d}\right)^{u+v-2} (|\mathbb{E}\Delta_{s_u}\Delta_{t_1}|) \\
&\leq \left(\frac{C\mu_n}{nh^d}\right)^{u+v} h^{2d}. \qquad (22)
\end{aligned}
$$

We then take the $\frac{d}{2d+2}$-power of (21) and the $\frac{d+2}{2d+2}$-power of (22)

$$
|\mathrm{Cov}(\Delta_{s_1}, \ldots \Delta_{s_u}, \Delta_{t_1}, \ldots \Delta_{t_v})| \leq h^d \left(\frac{C\mu_n}{nh^d}\right)^{u+v} ve^{-\frac{ad}{2d+2}(t_1-s_u)}.
$$

Applying Theorem 2.1 of [17] for the variables Δ_i, $i = 1, \ldots, n$ where

$$
K_n = \frac{C\mu_n}{n\sqrt{h^d}}, \; M_n = \frac{C\mu_n}{nh^d} \; \text{ and } A_n = \mathrm{Var}\left(\sum_{i=1}^{n} \Delta_i\right) = O\left(\frac{1}{nh^d}\right),
$$

it follows that

$$
\begin{aligned}
&\mathbb{P}\left(\left|\hat{\Psi}_N^*(x_k,t)) - \mathbb{E}\left[\hat{\Psi}_N^*(x_k,t))\right]\right| > \eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right) \\
&= \mathbb{P}\left(\left|\sum_{i=1}^{n}\Delta_i\right| > \eta\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right) \\
&\leq \exp\left\{-\frac{\eta^2\log n/(2n^{1-\gamma}h^d)}{\left(\operatorname{Var}\left(\sum_{i=1}^{n}\Delta_i\right) + C\mu_n(nh^d)^{-\frac{1}{3}}\left(\frac{\log n}{n^{1-\gamma}h^d}\right)^{\frac{5}{6}}\right)}\right\} \\
&\leq \exp\left\{-\frac{\eta^2\log n}{Cn^{-\gamma} + \mu_n n^{-\gamma/6}\left(\frac{\log^5 n}{nh^d}\right)^{\frac{1}{6}}}\right\} \\
&\leq C'\exp\left\{-C\eta^2\log n\right\}. \qquad (23)
\end{aligned}
$$

By using the fact that, $d_n z_n \leq n^{\zeta}$, where $\zeta = \beta + \frac{1}{2}$, we get

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq k\leq d_n}\max_{1\leq j\leq z_n}\max_{t\in\{t_j-l_n,t_j+l_n\}}\left|\hat{\Psi}_N^*(x,z) - \mathbb{E}\left[\hat{\Psi}_N^*(x,z)\right]\right| > \eta\sqrt{\frac{\log n}{nh^d}}\right) \\
&\leq C' n^{\beta+1/2-C\eta^2}.
\end{aligned}
$$

A suitable choice of η allows to finish the proof of (10).

- Secondly, concerning T_1 and T_3: The Lipschitz condition on the kernel K in (U6) allows to write directly, for all $x \in S$, and $\forall t \in [-\delta_0, \delta_0]$

$$
\begin{aligned}
\left|\hat{\Psi}_N^*(x,t) - \hat{\Psi}_N^*(x_{k(x)},t)\right| &= \frac{\mu_n}{nh^d}\left|\sum_{i=1}^{n}K_i(x) - \sum_{i=1}^{n}K_i(x_{k(x)})\right| \\
&\leq \frac{\mu_n}{h^{d+1}}\|x - x_{k(x)}\| \\
&\leq \frac{C\mu_n\tau_n}{h^{d+1}}.
\end{aligned}
$$

Since $\tau_n = O\left(n^{-\beta}\right)$, we have

$$
\frac{\tau_n}{h^{d+1}} = o\left(\sqrt{\frac{\log n}{nh^d}}\right) \qquad \text{a.co.} \qquad (24)
$$

Hence

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N^*(x,t)-\hat{\Psi}_N^*(x_{k(x)},t)\right| = O\left(\sqrt{\frac{\log n}{nh^d}}\right) \qquad \text{a.co} \tag{25}$$

and

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\left[\hat{\Psi}_N^*(x_{k(x)},t)\right]-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right| = O\left(\sqrt{\frac{\log n}{nh^d}}\right). \tag{26}$$

Now, we prove (8). For all $x \in S$ and all $t \in [-\delta_0, \delta_0]$, we have

$$\begin{aligned}
\left|\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right| &= \frac{1}{nh^d}\left|\mathbb{E}\left[\sum_{i=1}^n \psi(Y_i,t)\,\mathbb{1}_{\{|\psi(Y_i,t)|>\mu_n\}}K_i\right]\right| \\
&\le h^{-d}\mathbb{E}\left[|\psi(Y_1,t)|\,\mathbb{1}_{\{|\psi(Y_i,t)|>\mu_n\}}K_1\right] \\
&\le h^{-d}\mathbb{E}\left[\exp\left(|\psi(Y_1,t)|/4\right)\,\mathbb{1}_{\{|\psi(Y_i,t)|>\mu_n\}}K_1\right].
\end{aligned}$$

Using Holder's inequality, we get

$$\begin{aligned}
&\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right| \\
&\le h^{-d}\left(\mathbb{E}\left[\exp\left(|\psi(Y_1,t)|/2\right)\,\mathbb{1}_{\{|\psi(Y_i,t)|>\mu_n\}}\right]\right)^{\frac{1}{2}}\left(\mathbb{E}(K_1^2)\right)^{\frac{1}{2}} \\
&\le h^{-d}\exp\left(-\mu_n/4\right)\left(\mathbb{E}\left[\exp\left(|\psi(Y_1,t)|\right)\right]\right)^{\frac{1}{2}}\left(\mathbb{E}(K_1^2)\right)^{\frac{1}{2}} \\
&\le Ch^{\frac{-d}{2}}\exp\left(-\mu_n/4\right).
\end{aligned}$$

Since $\mu_n = n^{\gamma/6}$, we can then write

$$\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\mathbb{E}\left[\hat{\Psi}_N(x,t)\right]-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right| = o\left(\left(\frac{\log n}{n^{1-\gamma}h^d}\right)^{1/2}\right).$$

The last claimed result (9) is shown by using the Markov's inequality. Indeed, observe that, for all k, for all $t \in \mathcal{G}_n$ and for all $\epsilon > 0$

$$\begin{aligned}
&\mathbb{P}\left[\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N(x,t)-\hat{\Psi}_N^*(x,t)\right|>\epsilon\right] \\
&\quad = \mathbb{P}\left(\frac{1}{nh^d}\sum_{i=1}^n \psi(Y_i,t)\,\mathbb{1}_{|\psi(Y_i,t)|>\mu_n}K_i| > \epsilon\right)
\end{aligned}$$

$$
\begin{aligned}
&\le n\mathbb{P}\left(|\psi(Y_1,t)| > \mu_n\right)\\
&\le n\exp\left(-\mu_n\right)\mathbb{E}\left(\exp\left(|\psi(Y_1,t)|\right)\right)\\
&\le Cn\exp\left(-\mu_n\right).
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\sum_{n\ge 1}\mathbb{P}\left(\sup_{x\in S}\sup_{t\in[-\delta_0,\delta_0]}\left|\hat{\Psi}_N(x,t)-\hat{\Psi}_N^*(x,t)\right| > \epsilon_0\left(\sqrt{\frac{\log n}{n^{1-\gamma}h^d}}\right)\right)\\
&\quad\le C\sum_{n\ge 1} n\exp\left(-\mu_n\right). \qquad (27)
\end{aligned}
$$

Using the definition of μ_n completes the proof of (9) which in turn completes the proof of lemma.

Lemma 3 *Under hypotheses of Lemma 2, we have*

$$
\sup_{x\in S}\left|\mathbb{E}\left[\hat{\Psi}_D(x)\right]-\hat{\Psi}_D(x)\right| = O\left(\sqrt{\frac{\log n}{nh^d}}\right) \qquad \text{a.co.}
$$

Proof of Lemma 3 Similarly to Lemma 2, we use the compactness of S with respect the notations of the previous lemma and we write

$$
\begin{aligned}
\sup_{x\in S}\left|\hat{\Psi}_D(x)-\mathbb{E}\left[\hat{\Psi}_D(x)\right]\right| &\le \underbrace{\sup_{x\in S}\left|\hat{\Psi}_D(x)-\hat{\Psi}_D(x_k)\right|}_{T_1'} + \underbrace{\sup_{x\in S}\left|\hat{\Psi}_D(x_k)-\mathbb{E}\left[\hat{\Psi}_D(x_k)\right]\right|}_{T_2'}\\
&\quad + \underbrace{\sup_{x\in S}\left|\mathbb{E}\left[\hat{\Psi}_D(x_k)\right]-\mathbb{E}\left[\hat{\Psi}_D(x)\right]\right|}_{T_3'}.
\end{aligned}
$$

- For T_1' and T_3': the Lipschitz condition on the kernel K permit to write, for all $x \in S$

$$
\begin{aligned}
\left|\hat{\Psi}_D(x)-\hat{\Psi}_D(x_k)\right| &= \frac{1}{nh^d}\left|\sum_{i=1}^n K_i(x)-\sum_{i=1}^n K_i(x_k)\right|\\
&\le \frac{C}{h^{d+1}}\|x-x_k\|\\
&\le \frac{C\tau_n}{h^{d+1}}.
\end{aligned}
$$

The result follows directly, by the fact that $\frac{\tau_n}{h^{d+1}} = O\left(\sqrt{\frac{\log n}{nh^d}}\right)$. Thus

$$\sup_{x\in S}\left|\hat{\Psi}_D(x) - \hat{\Psi}_D(x_k)\right| = O\left(\sqrt{\frac{\log n}{nh^d}}\right) \qquad \text{a.co.} \tag{28}$$

and

$$\sup_{x\in S}\left|\mathbb{E}\left[\hat{\Psi}_D(x_k)\right] - \mathbb{E}\left[\hat{\Psi}_D(x)\right]\right| = O\left(\sqrt{\frac{\log n}{nh^d}}\right) \qquad \text{a.co.} \tag{29}$$

- For T_2', for all real $\eta > 0$, we have

$$\begin{aligned}\mathbb{P}\left(T_2' > \eta\sqrt{\frac{\log n}{nh^d}}\right) &\le \mathbb{P}\left(\max_{k\in\{1,\dots d_n\}}\left|\hat{\Psi}_D(x_k) - \mathbb{E}\left[\hat{\Psi}_D(x_k)\right]\right| > \eta\sqrt{\frac{\log n}{nh^d}}\right)\\ &\le d_n \max_{k\in\{1,\dots d_n\}}\mathbb{P}\left(\left|\hat{\Psi}_D(x_k) - \mathbb{E}\left[\hat{\Psi}_D(x_k)\right]\right| > \eta\sqrt{\frac{\log n}{nh^d}}\right).\end{aligned} \tag{30}$$

Proceeding as in Lemma 2 (replacing ψ by 1), we get,

$$\mathbb{P}\left(\sup_{x\in S}\left|\hat{\Psi}_D(x) - \mathbb{E}\left[\hat{\Psi}_D(x)\right]\right| > \eta\sqrt{\frac{\log n}{nh^d}}\right) \le n^{\beta - C\eta^2}.$$

A suitable choice of η allows to get

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\sup_{x\in S}\left|\mathbb{E}\left[\hat{\Psi}_D(x)\right] - \hat{\Psi}_D(x)\right| > \eta\sqrt{\frac{\log n}{nh^d}}\right) < \infty.$$

Lemma 4 *Under the hypotheses of Lemma 1, we have*

$$\sup_{x\in S}\left|f(x) - \mathbb{E}\left[\hat{\Psi}_D(x)\right]\right| = O(h).$$

Proof of Lemma 4 Using the same arguments as those in Lemma 1, it suffices to write

$$\begin{aligned}\mathbb{E}\left[\hat{\Psi}_D(x)\right] &= \frac{1}{h^d}\int_{\mathbb{R}^d} K\left(\frac{x-u}{h}\right) f(u)\,du\\ &= \int_{\mathbb{R}^d} f(x - hz)K(z)\,dz.\end{aligned}$$

Next, we obtain by analytical arguments

$$\sup_{x\in S}\left|f(x)-\mathbb{E}\left[\hat{\Psi}_D(x)\right]\right|=O(h).$$

Corollary 1 *Under the hypotheses of Lemma 3, we have*

$$\exists c>0 \qquad \sum_{n=1}^{\infty}\mathbb{P}\Big(\inf_{x\in S}\left|\hat{\Psi}_D(x)\right|\le c\Big)<\infty.$$

Proof of Corollary 1 We can write, $\forall x\in S$,

$$\inf_{x\in S}\hat{\Psi}_D(x)<\frac{\inf_{x\in S}f(x)}{2}\Rightarrow\sup_{x\in S}|f(x)-\hat{\Psi}_D(x)|>\frac{\inf_{x\in S}f(x)}{2}.$$

Then,

$$\mathbb{P}\left(\inf_{x\in S}|\hat{\Psi}_D(x)|\le\frac{\inf_{x\in S}f(x)}{2}\right)\le\mathbb{P}\left(\sup_{x\in S}|f(x)-\hat{\Psi}_D(x)|\ge\frac{\inf_{x\in S}f(x)}{2}\right).$$

The use of the results of Lemmas 3 and 4 completes the proof of the corollary.

Lemma 5 *Under the hypotheses of Theorem 1, $\hat{\theta}_x$ exists and is unique almost surely, for n large enough.*

Proof of Lemma 5 The strict monotony of ψ implies:

$$\Psi(x,\theta_x-\epsilon)\le\Psi(x,\theta_x)\le\Psi(x,\theta_x+\epsilon)$$

Lemmas 1–3 and Corollary 1 show that for all real fixed t.

$$\hat{\Psi}(x,t)-\Psi(x,t)\to 0 \qquad \text{a.co.}$$

So, for n sufficiently large, we have:

$$\hat{\Psi}(x,\theta_x-\epsilon)\le 0\le\hat{\Psi}(x,\theta_x+\epsilon) \qquad \text{a.co.}$$

As ψ and K are continuous functions, then $\hat{\Psi}(\dot{,}t)$ is continuous for all t. There exists $t_0=\hat{\theta}_x$ in some interval $[\theta_x-\epsilon,\theta_x+\epsilon]$ such that $\hat{\Psi}(x,\hat{\theta}_x)=0$. Finally, the unicity of $\hat{\theta}_x$ is a direct consequence of the strict monotony of ψ and the positivity of K.

Proof of Theorem 2 Similarly to Theorem 1, we give the proof for the case of a increasing ψ, decreasing case being obtained by considering $-\psi$. Considering this case, we define, for all $u \in \mathbb{R}$, $t = \theta_x + u\left[\hat{n}h^d\right]^{-1/2}\sigma(x, \theta_x)$. It is clear that, if $\hat{\Psi}_D(x) \neq 0$ we can write

$$\begin{aligned}
&\mathbb{P}\left\{\left(\frac{nh^d}{\sigma^2(x, \theta_x)}\right)^{1/2}\left(\widehat{\theta_x} - \theta_x\right) < u\right\} \\
&= \mathbb{P}\left\{\widehat{\theta_x} < \theta_x + u\left[\hat{n}h^d\right]^{-1/2}\sigma(x, \theta_x)\right\} \\
&= \mathbb{P}\left\{0 < \hat{\Psi}_N(x, t)\right\} \\
&= \mathbb{P}\left\{\mathbb{E}\left[\hat{\Psi}_N(x, t)\right] - \hat{\Psi}_N(x, t) < \mathbb{E}\left[\hat{\Psi}_N(x, t)\right]\right\}.
\end{aligned}$$

So, Theorem 2 is a consequence of the following intermediate results.

Lemma 6 *Under Hypotheses (H1), (U2), (U5) and (U7'), we have*

$$\mathbb{P}\left\{\left(\hat{\Psi}_D(x) = 0\right)\right\} \longrightarrow 0 \qquad \text{as} \quad n \longrightarrow \infty.$$

Proof of Lemma 6 Clearly, for all $\varepsilon < 1$, we have

$$\mathbb{P}\left\{\hat{\Psi}_D(x) = 0\right\} \leq \mathbb{P}\left\{\hat{\Psi}_D(x) \leq 1 - \varepsilon\right\} \leq \mathbb{P}\left\{|\hat{\Psi}_D(x) - 1| \geq \varepsilon\right\}.$$

So, it suffices to use Lemmas 3 and 4 to show that

$$\hat{\Psi}_D(x) - 1 \to 0 \quad \text{in probability.} \tag{31}$$

Lemma 7 *Under the hypotheses of Theorem 2, we have for any $x \in \mathscr{A}$*

$$\left(\frac{nh^d}{\left(\frac{\partial}{\partial t}\Psi(x, \theta_x)\right)^2\sigma^2(x, \theta_x)}\right)^{1/2}\left(\hat{\Psi}_N(x, t) - \mathbb{E}\left[\hat{\Psi}_N(x, t))\right]\right) \xrightarrow{\mathscr{D}} \mathscr{N}(0, 1) \quad \text{as } n \to \infty.$$

Proof of Lemma 7 Similarly to Lemma 2 we write, for a fixed $x \in S$

$$\begin{aligned}
\hat{\Psi}_N(x, t) - \mathbb{E}\left[\hat{\Psi}_N(x, t)\right] =& \hat{\Psi}_N(x, t) - \hat{\Psi}_N^*(x, t) \\
&+ \hat{\Psi}_N^*(x, t) - \mathbb{E}\left[\hat{\Psi}_N^*(x, t)\right] \\
&+ \mathbb{E}\left[\hat{\Psi}_N^*(x, t)\right] - \mathbb{E}\left[\hat{\Psi}_N(x, t)\right],
\end{aligned}$$

where $\hat{\Psi}_N^*(x,t)$ is given in Lemma 2. Once again we use the same arguments as in Lemma 2 to get

$$\left(\frac{nh^d}{\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2\sigma^2(x,\theta_x)}\right)^{1/2}\left|\hat{\Psi}_N(x,t)-\hat{\Psi}_N^*(x,t))\right| = o_p(1)$$

and

$$\left(\frac{nh^d}{\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2\sigma^2(x,\theta_x)}\right)^{1/2}\left|\mathbb{E}\left[\hat{\Psi}_N^*(x,t))\right]-\mathbb{E}\left[\hat{\Psi}_N(x,t))\right]\right| = o(1).$$

Then, it suffices to show the asymptotic normality of

$$\left(\frac{nh^d}{\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2\sigma^2(x,\theta_x)}\right)^{1/2}\left|\hat{\Psi}_N^*(x,t)-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right|.$$

For this we put

$$\psi^*(Y_i,t)=\psi_x(Y_i,t)1\!\!1_{|\psi(Y_i,t)|<\mu_n},\quad \Delta_i=\frac{1}{nh^d}\left(K_i\psi^*(Y_i,t)-\mathbb{E}\left[K_i\psi^*(Y_i,t)\right]\right),$$

$$Z_{ni}=\sqrt{nh^d}\Delta_i \quad \text{and} \quad S_n:=\sum_{i=1}^n Z_{ni}.$$

Therefore,

$$S_n=\sqrt{nh^d}\left(\hat{\Psi}_N^*(x,t)-\mathbb{E}\left[\hat{\Psi}_N^*(x,t)\right]\right).$$

Thus, our claimed result is, now

$$S_n \to \mathcal{N}(0,\sigma_1(x)), \tag{32}$$

where $\sigma_1^2(x)=\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2\sigma^2(x,\theta_x)$. To do that, we use the basic technique of [11] (pp. 228–231). Indeed, we consider $p=p_n$ and $q=q_n$, two sequences of natural numbers tending to ∞, such that

$$p=o\left(n^{1/2}\mu_n^{-1}h^{d/2}\right) \quad \text{and} \quad q=O(p^{1-\varsigma}), \quad \text{for a certain} \quad \varsigma\in(0,1)$$

and we split S_n into

$$S_n=T_n+T_n'+\zeta_k \ \text{ with } \ T_n=\sum_{j=1}^k \eta_j, \ \text{ and } \ T_n'=\sum_{j=1}^k \xi_j,$$

where

$$\eta_j := \sum_{i \in I_j} Z_{ni}, \quad \xi_j := \sum_{i \in J_j} Z_{ni}, \quad \zeta_k := \sum_{i=k(p+q)+1}^{n} Z_{ni},$$

with

$$I_j = \{ (j-1)(p+q)+1, \ldots, (j-1)(p+q)+p\},$$
$$J_j = \{ (j-1)(p+q)+p+1, \ldots, j(p+q)\}.$$

Observe that, for $k = \left[\frac{n}{p+q}\right]$, (where [.] stands for the integer part), we have $\frac{kq}{n} \to 0$, and $\frac{kp}{n} \to 1$, $\frac{q}{n} \to 0$, which imply that $\frac{p}{n} \to 0$ as $n \to \infty$. Now, our asymptotic result is based on

$$\mathbb{E}(T_n')^2 + \mathbb{E}(\zeta_k)^2 \to 0 \tag{33}$$

and

$$T_n \to \mathcal{N}(0, \sigma_1^2(x)). \tag{34}$$

For (33), we use the stationarity of variables to get

$$\mathbb{E}(T_n')^2 = k\mathrm{Var}(\xi_1) + 2 \sum_{1 \le i < j \le k} |\mathrm{Cov}(\xi_i, \xi_j)| \tag{35}$$

and

$$k\mathrm{Var}(\xi_1) \le qk\mathrm{Var}(Z_{n1}) + 2k \sum_{1 \le i < j \le q} |\mathrm{Cov}(Z_{ni}, Z_{nj})|. \tag{36}$$

The first term in the right-hand side of (36) can be deduced from (18) and the fact that $\frac{kq}{n} \to 0$. Indeed,

$$qk\mathrm{Var}(Z_{n1}) = h^d nkq\mathrm{Var}(\Delta_1) = O\left(\frac{kq}{n}\right) \to 0, \quad \text{as } n \to \infty. \tag{37}$$

The second term is

$$k \sum_{1 \le i < j \le q} |\mathrm{Cov}(Z_{ni}, Z_{nj})| = knh^d \sum_{1 \le i < j \le q} |\mathrm{Cov}(\Delta_i, \Delta_j)|$$

and similarly to (19) we show that

$$\sum_{1\le i<j\le q} |\mathrm{Cov}(\varDelta_i, \varDelta_j)| = o\left(\frac{q}{n^2h^d}\right).$$

Therefore,

$$k \sum_{1\le i<j\le q} |\mathrm{Cov}(Z_{ni}, Z_{nj})| = o\left(\frac{kq}{n}\right) \to 0, \quad \text{as } n \to \infty. \tag{38}$$

We use the stationarity to evaluate the second term in the right-hand side of (35)

$$\begin{aligned}
\sum_{1\le i<j\le k} |\mathrm{Cov}(\xi_i, \xi_j)| &= \sum_{l=1}^{k-1}(k-l)|\mathrm{Cov}(\xi_1, \xi_{l+1})| \\
&\le k\sum_{l=1}^{k-1} |\mathrm{Cov}(\xi_1, \xi_{l+1})| \\
&\le k\sum_{l=1}^{k-1} \sum_{(i,j)\in J_1\times J_{l+1}} \mathrm{Cov}(Z_{ni}, Z_{nj}).
\end{aligned}$$

It is clear that, for all $(i,j) \in J_1 \times J_j$, we have $|i-j| \ge p+1 > p$, then

$$\begin{aligned}
\sum_{1\le i<j\le k} |\mathrm{Cov}(\xi_i, \xi_j)| &\le k\frac{C\mu_n^2}{nh^{d+2}} \sum_{i=1}^{p} \sum_{j=2p+q+1, |i-j|>p}^{k(p+q)} \lambda_{i,j} \\
&\le \frac{Ckp\mu_n^2}{nh^{d+2}}\lambda_p \\
&\le \frac{Ckp\mu_n^2}{nh^{d+2}}e^{-ap} \to 0.
\end{aligned}$$

Finally, combining this last result with (36)–(38) we can write

$$\mathbb{E}(T_1')^2 \to 0 \;\text{ as } n \to \infty.$$

Since $(n - k(p+q)) \le p$, we have

$$\begin{aligned}
\mathbb{E}(\zeta_k)^2 &\le (n-k(p+q))\mathrm{Var}(Z_{n1}) + 2\sum_{1\le i<j\le n} |\mathrm{Cov}(Z_{ni}, Z_{nj})| \\
&\le p\mathrm{Var}(Z_{n1}) + 2\sum_{1\le i<j\le n} |\mathrm{Cov}(Z_{ni}, Z_{nj})|
\end{aligned}$$

$$\leq pnh^d \text{Var}(\Delta_1) + \underbrace{nh^d \sum_{1\leq i<j\leq n} |\text{Cov}(\Delta_i, \Delta_j)|}_{o(1)}$$

$$\leq \frac{Cp}{n} + o(1).$$

Then,

$$\mathbb{E}(\zeta_k)^2 \to 0, \quad \text{as } n \to \infty.$$

Proof of (34): it is based on

$$\left| \mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} \mathbb{E}\left(e^{it\eta_j}\right) \right| \to 0, \tag{39}$$

and

$$k\text{Var}(\eta_1) \to \sigma_1^2(x), \quad k\mathbb{E}(\eta_1^2 \mathbb{I}_{\{\eta_1 > \epsilon\sigma_1(x)\}}) \to 0. \tag{40}$$

Proof of (39):

$$\begin{aligned}
&\left| \mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} \mathbb{E}\left(e^{it\eta_j}\right) \right| \\
&\quad \leq \left| \mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \mathbb{E}\left(e^{it\sum_{j=1}^{k-1}\eta_j}\right)\mathbb{E}\left(e^{it\eta_k}\right) \right| \\
&\qquad + \left| \mathbb{E}\left(e^{it\sum_{j=1}^{k-1}\eta_j}\right) - \prod_{j=1}^{k-1} \mathbb{E}\left(e^{it\eta_j}\right) \right| \\
&\quad = \left| \text{Cov}\left(e^{it\sum_{j=1}^{k-1}\eta_j}, e^{it\eta_k}\right) \right| + \left| \mathbb{E}\left(e^{it\sum_{j=1}^{k-1}\eta_j}\right) - \prod_{j=1}^{k-1} \mathbb{E}\left(e^{it\eta_j}\right) \right|
\end{aligned} \tag{41}$$

and successively, we have

$$\begin{aligned}
\left| \mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k} \mathbb{E}\left(e^{it\eta_j}\right) \right| &\leq \left| \text{Cov}\left(e^{it\sum_{j=1}^{k-1}\eta_j}, e^{it\eta_k}\right) \right| \\
&\quad + \left| \text{Cov}\left(e^{it\sum_{j=1}^{k-2}\eta_j}, e^{it\eta_{k-1}}\right) \right| \\
&\quad + \cdots + \left| \text{Cov}\left(e^{it\eta_2}, e^{it\eta_1}\right) \right|.
\end{aligned} \tag{42}$$

Using the quasi-association property to write

$$\left|\operatorname{Cov}\left(e^{it\eta_2}, e^{it\eta_1}\right)\right| \leq \frac{Ct^2\mu_n^2}{nh^{d+2}} \sum_{i\in I_1}\sum_{j\in I_2} \lambda_{i,j}$$

and applying this inequality to each term on the right-hand side of (42), we obtain

$$\left|\mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k}\mathbb{E}\left(e^{it\eta_j}\right)\right|$$
$$\leq \frac{Ct^2\mu_n^2}{nh^{d+2}}\left[\sum_{i\in I_1}\sum_{j\in I_2}\lambda_{i,j} + \sum_{i\in I_1\cup I_2}\sum_{j\in I_3}\lambda_{i,j} + \cdots + \sum_{i\in I_1\cup\cdots\cup I_{k-1}}\sum_{j\in I_k}\lambda_{i,j}\right].$$

Observe that for every $2 \leq l \leq k-1$, $(i,j) \in I_l \times I_{l+1}$, we have $|i-j| \geq q+1 > q$, then

$$\sum_{i\in I_1\cup\cdots\cup I_{l-1}}\sum_{j\in I_l}\lambda_{i,j} \leq p\lambda_q.$$

Therefore, Inequality (41) becomes

$$\begin{aligned}\left|\mathbb{E}\left(e^{it\sum_{j=1}^{k}\eta_j}\right) - \prod_{j=1}^{k}\mathbb{E}\left(e^{it\eta_j}\right)\right| &\leq \frac{Ct^2\mu_n^2}{nh^{d+2}}kp\lambda_q \\ &\leq \frac{Ct^2\mu_n^2}{nh^{d+2}}kpe^{-aq} \to 0.\end{aligned}$$

Concerning (40) we use the same arguments as those in (35), to get

$$\begin{aligned}\lim_{n\to\infty} k\operatorname{Var}(\eta_1) &= \lim_{n\to\infty} kp\operatorname{Var}(Z_{n1}) \\ &= \lim_{n\to\infty} kpnh^d\operatorname{Var}(\Delta_1).\end{aligned}$$

On the other hand

$$\begin{aligned}\operatorname{Var}(\Delta_1) =& \frac{1}{n^2h^{2d}}\left\{\mathbb{E}\left[K^2\left(h^{-1}(x-X_i)\right)\psi_x^2(Y_i,t)\right]\right. \\ & \left. - \mathbb{E}\left[K^2\left(h^{-1}(x-X_i)\right)\psi_x^2(Y_i,t)\,\mathbb{1}_{|\psi(Y_i,t)|>\mu_n}\right]\right\} \\ & -\frac{1}{n^2}\left(\frac{1}{h^d}\mathbb{E}\left[K\left(h^{-1}(x-X_i)\right)\psi_x(Y_i,t)\,\mathbb{1}_{|\psi(Y_i,t)|<\mu_n}\right]\right)^2.\end{aligned}$$

Similarly to (9) and Lemma 1 we show that

$$\operatorname{Var}(\Delta_1) = \frac{\sigma_1^2(x)}{n^2h^d} + o\left(\frac{1}{n^2h^d}\right). \tag{43}$$

Hence

$$k\operatorname{Var}(\eta_1) = \frac{kp\sigma_1^2(x)}{n} + o\left(\frac{kp}{n}\right) \to \sigma_1^2(x).$$

For the second part of (40), we use the fact that $|\eta_1| \leq Cp|Z_{n1}| \leq \dfrac{C\mu_n p}{\sqrt{nh^d}}$, and Tchebychev's inequality to get

$$\begin{aligned} k\mathbb{E}(\eta_1^2 \mathbb{I}_{\{\eta_1>\epsilon\sigma_1(x)\}}) &\leq \frac{C\mu_n^2p^2k}{nh^d}\mathbb{P}(\eta_1 > \epsilon\sigma_1(x)) \\ &\leq \frac{C\mu_n^2p^2k}{nh^d}\frac{\operatorname{Var}(\eta_1)}{\epsilon^2\sigma_1^2(x)} \\ &= O\left(\frac{\mu_n^2p^2}{nh^d}\right) \end{aligned}$$

which completes the proof. ∎

Lemma 8 *Under Hypotheses (U1), (U5), (U7') and if the bandwidth parameter h satisfies $nh^{2+d} \to 0$ as $n \to \infty$, we have*

$$\left(\frac{nh^d}{\left(\frac{\partial}{\partial t}\Psi(x,\theta_x)\right)^2\sigma^2(x,\theta_x)}\right)^{1/2}\mathbb{E}\left[\hat{\Psi}_N(x,t)\right] = u + o(1), \text{ as } \quad n \to +\infty.$$

Proof of Lemma 8 By simple analytical arguments we write

$$\mathbb{E}[\hat{\Psi}_N(x,t)] = \int_{\mathbb{R}^d} H(x-hz,t)K(z)\,dz.$$

Next, we use a Taylor expansion of $H(x-hz,\theta_x + u\left[nh^d\right]^{-1/2}\sigma(x,\theta_x))$ to write

$$\mathbb{E}[\hat{\Psi}_N(x,t)] = u\left[nh^d\right]^{-1/2}\sigma(x,\theta_x))\Psi'(x,\theta(x)) + O(h).$$

The result is then a consequence of (U7'). ∎

References

1. Barlow, R.E., Proschan, F.: Statistical Theory of Reliability and Life Testing. Probability Models. Holt, Rinehart & Winston, New York (1981)
2. Boente, G., Fraiman, R.: Robust nonparametric regression estimation. J. Multivar. Anal. **29**, 180–198 (1989)
3. Boente, G., Fraiman, R.: Asymptotic distribution of robust estimators for nonparametric models from mixing processes. Ann. Stat. **18**, 891–906 (1990)
4. Boente, G., Ganzalez-Manteiga, W., Pérez-Gonzalez, A.: Robust nonparametric estimation with missing data. J. Stat. Plann. Inference **139**, 571–592 (2009)
5. Bulinski, A., Shabanovich, E.: Asymptotical behaviour of some functionals of positively and negatively dependent random fields (in Russian). Fundam. Appl. Math. **4**, 479–492 (1998)
6. Bulinski, A., Suquet, C.: Normal approximation for quasi-associated random fields. Stat. Probab. Lett. **54**, 215–226 (2001)
7. Collomb, G., Härdle, W.: Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. Stoch. Process. Appl. **23**, 77–89 (1986)
8. Dedecker, J., Doukhan, P., Lang, G., León, R.J.R., Louhichi, S., Prieur, C.: Weak Dependence: With Examples and Applications. Lecture Notes in Statistics. Springer, New York (2007)
9. Douge, L.: Théorèmes limites pour des variables quasi-associées hilbertiennes. Ann. Inst. Stat. Univ. Paris **54**, 51–60 (2010)
10. Doukhan, P., Lang, G., Surgailis, D., Tyssière, G.: Dependence in Probability and Statistics. Lecture Notes in Statistics. Springer, Berlin (2010)
11. Doob, J.L.: Stochastic Processes. Wiley, New York (1953)
12. Esary, J., Proschan, F., Walkup, D.: Association of random variables with applications. Ann. Math. Stat. **38**, 1466–1476 (1967)
13. Györfi, L., Härdle, W., Sarda, P., Vieu, P.: Nonparametric Curve Estimation from Time Series. Lecture Notes in Statistics, vol. 60. Springer, New York (1989)
14. Härdle, W., Tsybakov, A.: Robust nonparametric regression with simultaneous scale curve estimation. Ann. Stat. **16**, 120–135 (1988)
15. Huber, P.: Robust estimation of a location parameter. Ann. Math. Stat. **35**, 73–101 (1964)
16. Jong-Dev, K., Proschan, F.: Negative association of random variables, with applications. Ann. Stat. **11**, 286–295 (1983)
17. Kallabis, R.S., Neumann, M.H.: An exponential inequality under weak dependence. Bernoulli **12**, 333–350 (2006)
18. Laïb, N., Ould Saïd, E.: A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. Can. J. Stat. **28**, 817–828 (2000)
19. Masry, E.: Recursive probability density estimation for weakly dependent stationary processes IEEE. Trans. Inform. Theory **32**, 254–267 (1986)
20. Masry, E.: Multivariate probability density estimation for associated processes: strong consistency and rates. Stat. Probab. Lett. **58**, 205–219 (2002)
21. Matula, P.: A note on the almost sure convergence of sums of negatively dependent random variables. Stat. Probab. Lett. **15**, 209–213 (1992)
22. Robinson, P.M.: Robust nonparametric autoregression. In: Robust and Nonlinear Time Series Analysis (Heidelberg, 1983). Lectures Note in Statistics, vol. 26, pp. 247–255. Springer, New York (1983)
23. Roussas, G.G.: Kernel estimates under association: strong uniform consistency. Stat. Probab. Lett. **12**, 215–224 (1991)
24. Roussas, G.G.: Curve estimation in random fields of associated processes. J. Nonparametric Stat. **2**, 215–224 (1993)
25. Roussas, G.G.: Prediction under association. In: Limnios, N., Nikulin, M.S. (eds.) Recent Advances in Reliability Theory: Methodology, Practice and Inference. Birkhäuser, Boston (2000)
26. Shashkin, A.P.: Quasi-association for system of Gaussian vectors. Usp. Mat. Nauk (2002)

BlockShrink Wavelet Density Estimator in ϕ-Mixing Framework

Mohammed Badaoui and Noureddine Rhomari

Abstract We study the integrated L_2-risk, of a wavelet BlockShrink density estimator based on dependent observations. We prove that the BlockShrink estimator is adaptive in a class of Sobolev space with unknown regularity for uniformly mixing processes with arithmetically decreasing mixing coefficients.

1 Introduction

The functional estimation by the wavelet method has been intensively used, these last years, in various areas. The popularity of these methods comes from the ease of their implementation, their flexibility, ability to catch details, and for their high compression ratio. In the statistical literature, different types of wavelet estimators have been proposed. The performance of the first ones depended on the density's regularity. Later, adaptive procedures, as thresholding estimators, were developed to construct an estimate which does not depend on the explicit knowledge of this regularity. Those estimators are to make a fine selection of the coefficients estimators $\hat{\beta}_{jk}$ of the wavelets coefficient β_{jk} and several thresholding techniques, including local, global, and block thresholdings, have been developed.

Donoho and Johnstone developed the theory of thresholding in a general framework in the beginning of 1990s. They introduced two techniques of local thresholding: soft and hard thresholding. The word "local" means that individual coefficients independently of each other are subject to a possible thresholding.

M. Badaoui
Ecole Nationale des Sciences Appliquées (ENSA) de Khouribga, Université Hassan 1er, Settat, Morocco

LaMSD et URAC 04, Département de Mathématiques et Informatique, Faculté des Sciences, Université Mohamed 1er, Oujda, Morocco
e-mail: med_badaoui79@yahoo.fr; mohammed.badaoui@uhp.ac.ma

N. Rhomari (✉)
LaMSD et URAC 04, Département de Mathématiques et Informatique, Faculté des Sciences, Université Mohamed 1er, Oujda, Morocco
e-mail: nrhomari@yahoo.fr; n.rhomari@ump.ma

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_2

The idea of block thresholding was introduced by Efroimovich [9] as part of Fourier analysis. It has been adapted to the wavelet context analysis by Kerkyacharian et al. [13]. They proposed a global thresholding. Its principle, instead of keeping or deleting individual wavelet coefficients, one can also keep or delete a whole j-level of coefficients.

The first localized block thresholding estimators have been developed, in case of iid observations, by Hall et al. [10, 11], Cai [3–6], and Chesneau [7]. This last author studied an L_p version of the BlockShrink estimator given by Cai [3] and proved by simulations that BlockShrink estimator is better than the local and global ones. All these works have been developed in the independent framework. Tribouley and Viennet [20] explored the global thresholding method in the case of β-mixing processes. But to our knowledge, the BlockShrink estimator for the density model has not been studied for dependent processes.

The aim of this work is to extend some results, of BlockShrink estimator, to dependent processes in L_2-norm for the density estimation. The function density f is supposed to belong to the Sobolev space $\mathbf{H}^s$ with compact support. We consider the ϕ-mixing's processes [12] and we study the L_2-error convergence for BlockShrink estimator f_n. We show that BlockShrink estimator is consistent with an optimal rate, under certain conditions on wavelet basis and mixing coefficients. Precisely we give for arithmetically decreasing ϕ-mixing processes an upper bound of the L_2-mean error

$$\mathbb{E}\|f - f_n\|_{L_2}^2 = Cn^{-\frac{2s}{1+2s}} .$$

The rest of the paper is organized as follows: Sect. 2 describes briefly the wavelet basis and the Sobolev space. Section 3 introduces the BlockShrink estimator and presents its asymptotic properties under ϕ-mixing conditions. Section 4 contains proofs of various results.

2 Wavelets and Sobolev Space

In this section, we give a brief definition of wavelet decomposition and we refer to Meyer [16] and Mallat [14, 15] for more details. Without loss of generality, we suppose that f has $[0, 1]$ as support and consider a wavelet basis on the interval $[0, 1]$ of the form

$$\zeta = \{\varphi_{\tau k},\ \tau \geq 0, k = 0, \ldots, 2^j - 1; \psi_{jk},\ j \geq \tau; k = 0, \ldots, 2^j - 1\} .$$

In general, $\varphi_{jk}(x)$ and $\psi_{jk}(x)$ are "periodic" or "boundary adjusted" dilation and translation of a "father" wavelet φ and a "mother" wavelet ψ, respectively. This last function is supposed to be N-regular (cf. Meyer [16]).

For the sake of simplicity, we set $\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k)$ and $\psi_{jk} = 2^{j/2}\psi(2^j x - k)$, where φ and ψ are father and mother wavelet functions, respectively. Let, now, τ be a large enough integer. For any $j_0 \geq \tau$, a function f in $L_2([0,1])$ can be expanded in a wavelet series as

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0 k}\varphi_{j_0 k} + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} \ , \tag{1}$$

where the wavelet coefficients are defined by

$$\alpha_{j_0 k} = \int f(x)\varphi_{j_0 k}(x)dx \quad \text{et} \qquad \beta_{jk} = \int f(x)\psi_{jk}(x)dx.$$

Now, let us give a definition of Sobolev space, the main function spaces used in our study. To unify notations we write $\beta_{\tau-1,k} = \alpha_{\tau k}$. We define the Sobolev space $\mathbf{H}^s$ as being the functions space f whose the associated wavelet coefficients β_{jk}, satisfy

$$\left(\sum_{j \geq \tau-1} 2^{2js} \left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 \right) \right)^{1/2} < \infty,$$

where $s \in]0, N+1[$ with N denotes the wavelet regularity (cf. Meyer [16]).

3 BlockShrink Estimator

Let $X_1, \ldots, X_n$, n observations from a strictly stationary process, with unknown density f, with respect to the Lebesgue measure on $\mathbf{R}$, which is supposed to be in $L_2([0,1])$.

Then f admits a representation in (1) with $\alpha_{j_0 k} = \mathbb{E}(\varphi_{j_0 k}(X))$ and $\beta_{jk} = \mathbb{E}(\psi_{jk}(X))$, where X is a random variable having as probability density f.

We define the nonlinear BlockShrink estimator as proposed by Cai [3] by

$$f_n^b = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0 k}\varphi_{j_0 k} + \sum_{j=j_0}^{j_1} \sum_{K \in \mathscr{A}_j} \sum_{k \in \mathscr{B}_{jK}} \hat{\beta}_{jk} I_{\{\hat{b}_{jK} \geq \lambda\}} \psi_{jk}, \tag{2}$$

where j_0 and j_1 are two integers to be precised below such that: j_0 is an integer chosen so that the linear variance term will not contribute to the overall error in the same spirit that the integer j_1 is chosen to make the bias term negligible in the overall error. We assume that $2^{j_1} \asymp n^{\frac{1}{2}}$ and $2^{j_0} \asymp (\log n)^{\varepsilon}$ with $\varepsilon > 2$. And for

$j \in \{j_0, \ldots, j_1\}$, we split the set $\{0, \ldots, 2^j - 1\}$ into blocks of length l_j. We define the sets:

$$\mathscr{A}_j = \{1, \ldots, 2^j l_j^{-1}\},$$
$$\mathscr{B}_{jK} = \{k \in \{0, \ldots, 2^j - 1\}; (K-1)l_j \le k \le Kl_j - 1\},$$

for $K \in \mathscr{A}_j$, where l_j is an increasing sequence in j such that $l_{j_0} \asymp (\log n)^{\varepsilon}$, and we get

$$\hat{b}_{jK} = \left(l_j^{-1} \sum_{k \in \mathscr{B}_{jK}} |\hat{\beta}_{jk}|^2 \right)^{\frac{1}{2}}$$

and λ is some threshold parameter, depending on n, to be defined below.

The choice of block size l_j and the threshold λ play a crucial role in the performance of the resulting estimator. Between the two parameters l_j and λ, the block l_j size is more important. It plays an important role similar to that of the bandwidth in the kernel estimator. To achieve the optimal adaptability, the block size must be at least of order $\log n$, see, for example, Hall et al. [10].

Before announcing our result, we recall the uniformly mixing dependance we work with. Let $(\Omega, \mathscr{A}, P)$ be a probability space and let $\mathscr{F}$ and $\mathscr{G}$ be two sub σ-fields of $\mathscr{A}$. In order to estimate the dependence between $\mathscr{F}$ and $\mathscr{G}$ we use the uniform mixing (or ϕ-mixing) coefficient [12]

$$\phi(\mathscr{F}, \mathscr{G}) = \sup\{|\mathbb{P}(B) - \mathbb{P}(B/A)|;\ B \in \mathscr{G}, A \in \mathscr{F} \text{ and } \mathbb{P}(A) > 0\}.$$

When $\mathscr{F} = \sigma(X)$ and $\mathscr{G} = \sigma(Y)$ are sigma algebras generated, respectively, by random variables X and Y, we write $\phi(\mathscr{F}, \mathscr{G}) = \phi(X, Y)$.

Remark that, this measure of dependence is not symmetric with respect to $\mathscr{F}$ and $\mathscr{G}$. Therefore we can define an other coefficient by

$$\phi_{\text{rev}}(\mathscr{F}, \mathscr{G}) = \phi(\mathscr{G}, \mathscr{F}).$$

A stationary process $\mathbf{X} = (X_j, j \in \mathbf{Z})$ is said to be ϕ-mixing if

$$\phi(l) = \phi(\mathscr{F}_{-\infty}^{0}, \mathscr{F}_{l}^{+\infty}) \to 0 \text{ when } l \to \infty,$$

where $\mathscr{F}_i^k$ is the sigma algebra generated by the random variables $\{X_j,\ i \le j \le k\}$. In the same way, we can also define a ϕ_{rev}-mixing processes.

The sequence of the mixing coefficients $(\phi(l))_{l \ge 0}$ decrease arithmetically if there exists $\theta > 0$ and a constant $c > 0$ such that $\phi(l) \le cl^{-(\theta+1)}$; and we say $\mathbf{X}$ is an arithmetically ϕ-mixing process.

We now give Theorem 1 providing an upper bound of the L_2 error of the BlockShrink estimator defined by (2).

Theorem 1 *Assume that f belongs to the class $\mathscr{F}(M_1, M_2, A) = \{f \in \mathbf{H}^s, supp(f) \subset [A, -A], \|f\|_{\mathbf{H}^s} \le M_1, \|f\|_\infty \le M_2\}$, with $1/2 < s < N+1$, and the mixing coefficients decrease arithmetically with $\theta > 6 - \frac{4}{1+2(N+1)}$, then for*

$\lambda = 4\|\psi\|_{L_1}\|\psi\|_\infty\|f\|_\infty(\sum_{l\ge 0}\phi(l))/n^{1/2}$ there exists a positive constant C such that

$$\mathbb{E}\|f - f_n^b\|_{L_2}^2 \le Cn^{-\frac{2s}{1+2s}}.$$

Corollary 1 *Under the same conditions of Theorem 1, the estimator f_n^b is adaptive in the class $\{\mathscr{F}(M_1, M_2, A), 1/2 < s < N+1, A > 0, M_1 > 0, M_2 > 0\}$.*

Proofs of our main theorem are based on the following preliminary results, which are important in themselves.

Proposition 1 *Let us set $\hat{\beta}_{j_0-1k} = \hat{\alpha}_{j_0k}$, under the summability $\sum_{l\ge 0}\phi(l) \le B < \infty$, we have for all $j \in \{j_0 - 1, \ldots, j_1\}$*

$$\mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^2 \le 4B\|\psi\|_{L_1}\|\psi\|_\infty\|f\|_\infty\frac{1}{n}.$$

Proposition 2 *Assume that there exists $\theta > 0$ and a constant $c > 0$ such that $\phi(l) \le cl^{-(\theta+1)}$, then there exists two positive constants C and K_2 such that, for all j*

$$\mathbb{P}\left(\left(l_j^{-1}\sum_{k\in\mathscr{B}_{jK}}|\hat{\beta}_{jk} - \beta_{jk}|^2\right)^{1/2} \ge 2K_2/n^{1/2}\right) \le Cn^{-\frac{\theta}{2}}\left(\frac{2^j}{l_j}\right)^{\frac{1}{2}(2+\theta)}(\log n)^{\theta+1},$$

where $K_2 = (4B\,\|\psi\|_{L_1}\|\psi\|_\infty\,\|f\|_\infty)^{\frac{1}{2}}$ and $B = \sum_{l\ge 0}\phi(l)$.

4 Derivations

In what follows, we use the symbol C as a generic positive constant, independent of n, which may take different values at different places.

Let us first prove Propositions 1 and 2. We are inspired by techniques developed in Tribouley and Viennet [20].

Proof of Proposition 1 It is easy to have

$$\mathbb{E}\left(|\hat{\beta}_{jk}-\beta_{jk}|^2\right) = n^{-2}\mathbb{E}\left(\left|\sum_{i=1}^{n}\left(\psi_{jk}(X_i)-\mathbb{E}\left(\psi_{jk}\right)\right)\right|^2\right)$$

$$\leq\ 2\,n^{-2}\sum_{i=0}^{n}(n-i)\left|\mathrm{Cov}\left(\psi_{jk}(X_0),\psi_{jk}(X_i)\right)\right|.$$

From, Rio [17, Theorem 1.4, p. 27], we deduce

$$\left|\mathrm{Cov}\left(\psi_{jk}(X_0),\psi_{jk}(X_i)\right)\right|$$
$$\leq 2\phi^{\frac{1}{p}}(\psi_{jk}(X_i),\psi_{jk}(X_0))\phi^{\frac{1}{q}}(\psi_{jk}(X_0),\psi_{jk}(X_i))\mathbb{E}\left(\|\psi_{jk}(X_0)\|_{L_q}\right)\mathbb{E}\left(\|\psi_{jk}(X_0)\|_{L_p}\right).$$

Then, for $q=1$ and $p=\infty$ we have

$$\mathbb{E}\left(\|\psi_{jk}(X_0)\|_{L_1}\right)\leq 2^{-\frac{j}{2}}\|\psi\|_{L_1}\|f\|_\infty \text{ and } \mathbb{E}\left(\|\psi_{jk}(X_0)\|_\infty\right)\leq 2^{\frac{j}{2}}\|\psi\|_\infty,$$

and finally, we obtain the desired inequality

$$\mathbb{E}\left(|\hat{\beta}_{jk}-\beta_{jk}|^2\right)\leq 4\,n^{-1}\sum_{i=0}^{n}\phi(\psi_{jk}(X_0),\psi_{jk}(X_i))\mathbb{E}\left(\|\psi_{jk}(X_0)\|_{L_1}\right)\|\psi_{jk}(X_0)\|_\infty$$

$$\leq 4\,n^{-1}\|\psi\|_{L_1}\|\psi\|_\infty\|f\|_\infty\sum_{i=0}^{n}\phi(i).$$

■

Proof of Proposition 2 To prove Proposition 2, we need the following theorem which gives an exponential type inequality of block of $(\hat{\beta}_{jk})$ coefficients.

Theorem 2 *Let $(X_i)_{i\geq 0}$ be a strictly stationary uniformly mixing process such that its sequence of mixing coefficients $(\phi(l))_{l\geq 0}$ satisfies the summability $\sum_{l\geq 0}\phi(l)\leq B<\infty$. We suppose that X_i has a density f with respect to the Lebesgue measure and that f is uniformly bounded. Then, there exists a positive constant K_1 depending on $\|f\|_\infty$ and K_2 such that for any integer $q\leq n$, $\lambda_1>0$ and $\lambda_2>0$ we have*

$$\mathbb{P}\left(\left(\sum_{k\in\mathscr{B}_{jK}}|\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{\frac{1}{2}}\geq\lambda_1+K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}+\lambda_2\right)$$

$$\leq\exp\left(-K_1 n\left(\lambda_1^2 l_j^{-1}\wedge\frac{\lambda_1}{q2^{\frac{j}{2}}}\wedge\frac{\lambda_1^2\sqrt{n}}{q l_j^{\frac{1}{2}}2^{\frac{j}{2}}}\right)\right)+\frac{2C}{\lambda_2}2^{\frac{j}{2}}\phi(q)$$

with $K_2=(4B\,\|\psi\|_{L_1}\|\psi\|_\infty\,\|f\|_\infty)^{\frac{1}{2}}$ and $x\wedge y=\min(x,y)$.

The proof of this theorem is postponed to the end.

Let us return to the proof of Proposition 2. Applying this Theorem 2 with: $\lambda_1 = \lambda_2 = \frac{1}{2}K_2 l_j^{\frac{1}{2}} \frac{1}{\sqrt{n}}$, we get: for all $q \le n$

$$\mathbb{P}\left(\left(\sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{\frac{1}{2}} \ge 2K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}\right)$$
$$\le \exp\left(-K_1 n\left(\left(\frac{1}{2}K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}\right)^2 l_j^{-1} \wedge \frac{\frac{1}{2}K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}}{q2^{\frac{j}{2}}} \wedge \frac{\left(\frac{1}{2}K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}\right)^2\sqrt{n}}{q l_j^{\frac{1}{2}} 2^{\frac{j}{2}}}\right)\right)$$
$$+\frac{2C}{\frac{1}{2}K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}} 2^{\frac{j}{2}}\phi(q)$$
$$\le \exp\left(-K_1 n\left(\frac{1}{n}\wedge \frac{l_j^{\frac{1}{2}}}{q\sqrt{n}2^{\frac{j}{2}}} \wedge \frac{l_j^{\frac{1}{2}}}{q\sqrt{n}2^{\frac{j}{2}}}\right)\right) + C\sqrt{n} l_j^{-\frac{1}{2}} 2^{\frac{j}{2}}\phi(q)$$
$$\le \exp\left(-K_1\frac{1}{q} l_j^{\frac{1}{2}} 2^{-\frac{j}{2}}\sqrt{n}\right) + C\sqrt{n} l_j^{-\frac{1}{2}} 2^{\frac{j}{2}}\phi(q).$$

Now for $q = [l_j^{\frac{1}{2}} 2^{-\frac{j}{2}}\sqrt{n}(\eta\log n)^{-1}]$ with η a positive constant to be precised, we have

$$\mathbb{P}\left(\left(\sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{\frac{1}{2}} \ge 2K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}\right)$$
$$\le \exp(-K_1\eta\log n) + C\sqrt{n} l_j^{-\frac{1}{2}} 2^{\frac{j}{2}}\phi\left(l_j^{\frac{1}{2}} 2^{-\frac{j}{2}}\sqrt{n}(\eta\log n)^{-1}\right).$$

So for a good choice of η, the exponential term will be less than the mixing part (ie. the last term in the above inequality, thanks to the arithmetically decrease of ϕ), which gives

$$\mathbb{P}\left(\left(\sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{\frac{1}{2}} \ge 2K_2 l_j^{\frac{1}{2}}\frac{1}{\sqrt{n}}\right) \le C\sqrt{n} l_j^{-\frac{1}{2}} 2^{\frac{j}{2}}\phi\left(l_j^{\frac{1}{2}} 2^{-\frac{j}{2}}\sqrt{n}(\eta\log n)^{-1}\right).$$

As the mixing is arithmetic, $\phi(l) \le c\, l^{-(\theta+1)}$, we get

$$\mathbb{P}\left(\left(\sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2\right)^{\frac{1}{2}} \ge 2\frac{K_2 l_j^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right) \le C\sqrt{n} l_j^{-\frac{1}{2}} 2^{\frac{j}{2}} \left(l_j^{\frac{1}{2}} 2^{-\frac{j}{2}} \sqrt{n}(\eta \log n)^{-1}\right)^{-(\theta+1)}$$

$$\le Cn^{-\frac{\theta}{2}} \left(\frac{2^j}{l_j}\right)^{\frac{1}{2}(2+\theta)} (\log n)^{(\theta+1)}.$$

■

Proof of Theorem 1 The proof uses techniques from conventional calculations. In fact, recall that

$$f = \sum_{k=0}^{2^{j_0}-1} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\ge j_0}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}$$

$$f_n = \sum_{k=0}^{2^{j_0}-1} \hat{\alpha}_{j_0k}\varphi_{j_0k} + \sum_{j=j_0}^{j_1}\sum_{K\in\mathscr{A}_j}\sum_{k\in\mathscr{B}_{jK}} \hat{\beta}_{jk} I_{\{\hat{b}_{jK}\ge\lambda\}}\psi_{jk}$$

with $\lambda/2 = 2K_2/n^{1/2}$. It is obvious that

$$\mathbb{E}\|f-f_n\|_2^2 \le 3\left(\mathbb{E}\left\|\sum_{k=0}^{2^{j_0}-1}(\hat{\alpha}_{j_0k}-\alpha_{j_0k})\varphi_{j_0k}\right\|_{L_2}^2\right.$$
$$\left. + \mathbb{E}\left\|\sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1}(\hat{\beta}_{jk} I_{\{\hat{b}_{jK}\ge\lambda\}} - \beta_{jk})\psi_{jk}\right\|_{L_2}^2 + \left\|\sum_{j\ge j_1+1}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_{L_2}^2\right). \tag{3}$$

1. **Bias term:** Applying the Hölder's inequality and the fact that $f \in \mathscr{F}(M_1, M_2, B)$ we obtain

$$\left\|\sum_{j\ge j_1+1}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_{L_2}^2 = \left(\sum_{j\ge j_1+1}\left(\sum_{k=0}^{2^j-1}|\beta_{jk}|^2\right)^{1/2}\right)^2$$

$$= \left(\sum_{j\ge j_1+1} 2^{js}\left(\sum_{k=0}^{2^j-1}|\beta_{jk}|^2\right)^{1/2} 2^{-js}\right)^2$$

$$\leq \left(\left(\sum_{j\geq j_1+1} 2^{2js} \left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 \right) \right)^{1/2} \left(\sum_{j\geq j_1+1} 2^{-2js} \right)^{1/2} \right)^2$$

$$\leq 2^{-2j_1 s} \leq Cn^{-s} \leq Cn^{-\frac{2s}{1+2s}}. \tag{4}$$

2. **Linear stochastic term:** Using the Proposition 1, we obtain

$$\mathbb{E} \left\| \sum_{k=0}^{2^{j_0}-1} (\hat{\alpha}_{j_0 k} - \alpha_{j_0 k}) \varphi_{j_0 k} \right\|_{L_2}^2 = C \sum_{k=0}^{2^{j_0}-1} \mathbb{E} \left| \hat{\alpha}_{j_0 k} - \alpha_{j_0 k} \right|^2$$
$$\leq C \left(\frac{2^{j_0}}{n} \right) \leq Cn^{-\frac{2s}{1+2s}}. \tag{5}$$

3. **Nonlinear stochastic term:** Using techniques similar to those used in the proofs of Theorems 2 and 3 of Chesneau [8], we give

$$\mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\hat{\beta}_{jk} I_{\{\hat{b}_{jK} \geq \lambda\}} - \beta_{jk} \right) \psi_{jk} \right\|_{L_2}^2$$
$$\leq 2 \left(\mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk} I_{\{\hat{b}_{jK} < \lambda\}} \psi_{jk} \right\|_{L_2}^2 + \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\hat{\beta}_{jk} - \beta_{jk} \right) I_{\{\hat{b}_{jK} \geq \lambda\}} \psi_{jk} \right\|_{L_2}^2 \right)$$
$$= 2(G_2 + G_3).$$

(a) The upper bound for G_2. We have $G_2 \leq 2\,(G_{21} + G_{22})$, where

$$G_{21} = \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk} I_{\{\hat{b}_{jK} < \lambda\}} I_{\{b_{jK} \leq 2\lambda\}} \psi_{jk} \right\|_{L_2}^2$$

$$G_{22} = \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \beta_{jk} I_{\{\hat{b}_{jK} < \lambda\}} I_{\{b_{jK} > 2\lambda\}} \psi_{jk} \right\|_{L_2}^2 .$$

Notice that the l_2 Minkowski's inequality yields

$$I_{\{\hat{b}_{jK} < \lambda\}} I_{\{b_{jK} > 2\lambda\}} \leq I_{\{|\hat{b}_{jK} - b_{jK}| \geq \lambda\}}$$
$$\leq I_{\left\{ \left(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk} - \beta_{jk}|^2 \right)^{1/2} \geq \lambda \right\}}. \tag{6}$$

- For bounding the term G_{22} we will use the Large deviation established in Proposition 2. Using the inequality (6) and the Proposition 2 we obtain

$$
\begin{aligned}
G_{22} &\leq \mathbb{E}\int \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 I_{\{\hat{b}_{jK}<\lambda\}} I_{\{b_{jK}>2\lambda\}} |\psi_{jk}(x)|^2 dx \\
&\leq \int \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 |\psi_{jk}(x)|^2 \mathbb{E}\left[I_{\{\hat{b}_{jK}<\lambda\}} I_{\{b_{jK}>2\lambda\}} \right] dx \\
&\leq \int \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 |\psi_{jk}(x)|^2 P\left(\left(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2 \right)^{1/2} \geq \lambda \right) dx \\
&\leq \left\| \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} \left[P\left(\left(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2 \right)^{1/2} \geq \lambda \right) \right]^{1/2} \right\|_{L_2}^2 \\
&\leq C n^{-\frac{\theta}{2}} l_{j_1}^{-\frac{1}{2}(2+\theta)} 2^{\frac{j_1}{2}(2+\theta)} (\log n)^{(\theta+1)} \left\| \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} \right\|_{L_2}^2 \\
&\leq C n^{-\frac{\theta}{2}} \left(\frac{2^{j_1}}{l_{j_1}} \right)^{\frac{1}{2}(2+\theta)} (\log n)^{(\theta+1)} \|f\|_{L_2}^2 .
\end{aligned}
$$

This quantity is bounded by $n^{-\frac{2s}{1+2s}}$ for $2^{j_1} \asymp n^{\frac{1}{2}}$ with $\theta > 6 - \frac{4}{1+2(N+1)}$.
We then obtain for $\theta > 6 - \frac{4}{1+2(N+1)}$ that

$$
G_{22} \leq C n^{-\frac{2s}{1+2s}}. \tag{7}
$$

- The upper bound for G_{21}. Let j_s the integer belonging to $[j_0, j_1]$ such that j_s is the optimal level of truncation, when the regularity s of the density f is known (i.e $2^{j_s} \asymp n^{\frac{1}{1+2s}}$). So using an elementary inequality of convexity, we have the following decomposition:

$$
\begin{aligned}
G_{21} &\leq C \left\| \sum_{j=j_0}^{j_1}\sum_{k=0}^{2^j-1} \beta_{jk} I_{\{b_{jK}\leq 2\lambda\}} \psi_{jk} \right\|_{L_2}^2 \\
&\leq C \left(\left\| \sum_{j=j_0}^{j_s}\sum_{k=0}^{2^j-1} \beta_{jk} I_{\{b_{jK}\leq 2\lambda\}} \psi_{jk} \right\|_{L_2}^2 + \left\| \sum_{j=j_s+1}^{j_1}\sum_{k=0}^{2^j-1} \beta_{jk} I_{\{b_{jK}\leq 2\lambda\}} \psi_{jk} \right\|_{L_2}^2 \right)
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\left(\left\|\sum_{j=j_0}^{j_s}\sum_{k=0}^{2^j-1}\beta_{jk}I_{\{b_{jK}\leq 2\lambda\}}\psi_{jk}\right\|_{L_2}^2+\left\|\sum_{j=j_s+1}^{j_1}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_{L_2}^2\right)\\
&= C(S_1+S_2). \qquad (8)
\end{aligned}
$$

– The upper bound for S_1. Minkowski's inequality gives the following bound:

$$
\begin{aligned}
S_1 &= \left\|\sum_{j=j_0}^{j_s}\sum_{k=0}^{2^j-1}\beta_{jk}I_{\{b_{jK}\leq 2\lambda\}}\psi_{jk}\right\|_{L_2}^2\\
&= \left(\sum_{j=j_0}^{j_s}\left(\sum_{k=0}^{2^j-1}\left|\beta_{jk}I_{\{b_{jK}\leq 2\lambda\}}\right|^2\right)^{1/2}\right)^2.
\end{aligned}
$$

Thus

$$
\begin{aligned}
S_1 &= C\left(\sum_{j=j_0}^{j_s}\left(\sum_{K\in\mathscr{A}_j}\sum_{k\in\mathscr{B}_{jK}}|\beta_{jk}|^2 I_{\left\{\left(l_j^{-1}\sum_{k\in\mathscr{B}_{jK}}|\beta_{jk}|^2\right)^{1/2}\leq 2\lambda\right\}}\right)^{1/2}\right)^2\\
&\leq C\left[\sum_{j=j_0}^{j_s}\left(\mathrm{card}\mathscr{A}_j\, l_j\, \lambda^2\right)^{1/2}\right]^2\\
&\leq C\left[\sum_{j=j_0}^{j_s}2^{j/2}\,\lambda\right]^2\\
&\leq C2^{j_s}\lambda^2.
\end{aligned}
$$

– The upper bound for S_2. The Minkowski's inequality, the basis orthonormality and the inclusion $\mathbf{H}^s\subseteq\mathbf{B}^s_{2,\infty}$ imply that

$$
\begin{aligned}
S_2 &= \left\|\sum_{j=j_s+1}^{j_1}\sum_{k=0}^{2^j-1}\beta_{jk}\psi_{jk}\right\|_{L_2}^2=\left(\sum_{j=j_s+1}^{j_1}\left(\sum_{k=0}^{2^j-1}|\beta_{jk}|^2\right)^{1/2}\right)^2\\
&= \left(\sum_{j=j_s+1}^{j_1}2^{js}\left(\sum_{k=0}^{2^j-1}|\beta_{jk}|^2\right)^{1/2}2^{-js}\right)^2.
\end{aligned}
$$

Then, we obtain

$$
\begin{aligned}
S_2 &\leq \left(\sup_{j \geq j_0} 2^{js} \left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 \right)^{1/2} \sum_{j=j_s+1}^{j_1} 2^{-js} \right)^2 \\
&\leq \left(\|f\|_{\mathbf{B}^s_{2,\infty}} \sum_{j=j_s+1}^{j_1} 2^{-js} \right)^2 \\
&\leq C \left(2^{-j_s s} \sum_{j=1}^{j_1} 2^{-js} \right)^2 \\
&\leq C 2^{-2j_s s}.
\end{aligned}
$$

Putting the upper bounds of S_1 and S_2 together in inequality (8), it yields

$$
G_{21} \leq C \left(2^{-2j_s s} + 2^{j_s} \lambda^2 \right).
$$

Since $\lambda = 4K_2/n^{1/2}$ and $2^{j_s} \asymp n^{\frac{1}{1+2s}}$, we conclude that

$$
G_{21} \leq C n^{-\frac{2s}{1+2s}}. \tag{9}
$$

(b) For the term G_3, we have

$$
\begin{aligned}
G_3 &= \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\hat{\beta}_{jk} - \beta_{jk} \right) I_{\{\hat{b}_{jK} \geq \lambda\}} \psi_{jk} \right\|_{L_2}^2 \\
&\leq 2 \left(\mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (\hat{\beta}_{jk} - \beta_{jk}) I_{\{\hat{b}_{jK} \geq \lambda\}} I_{\{b_{jK} \leq \lambda/2\}} \psi_{jk} \right\|_{L_2}^2 \right. \\
&\quad \left. + \mathbb{E} \left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (\hat{\beta}_{jk} - \beta_{jk}) I_{\{\hat{b}_{jK} \geq \lambda\}} I_{\{b_{jK} > \lambda/2\}} \psi_{jk} \right\|_{L_2}^2 \right) \\
&= 2 \left(G_{31} + G_{32} \right).
\end{aligned} \tag{10}
$$

- To bound the term G_{31} we will use the Large deviation established in Proposition 2. Using l_2 Minkowski's inequality, Hölder's inequality and

Jensen's inequality we obtain

$$
\begin{aligned}
G_{31} &= \mathbb{E}\left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\hat{\beta}_{jk} - \beta_{jk}\right) I_{\left\{\hat{b}_{jK} \geq \lambda\right\}} I_{\left\{b_{jK} < \lambda/2\right\}} \psi_{jk} \right\|_{L_2}^2 \\
&\leq \mathbb{E} \int \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} |\hat{\beta}_{jk} - \beta_{jk}|^2 I_{\left\{(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2)^{1/2} \geq \lambda/2\right\}} |\psi_{jk}(x)|^2 dx \\
&\leq \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \int \mathbb{E}\left(|\hat{\beta}_{jk} - \beta_{jk}|^2 I_{\left\{(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2)^{1/2} \geq \lambda/2\right\}} \right) |\psi_{jk}(x)|^2 dx \\
&\leq \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \left(\mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^4\right)^{\frac{1}{2}} \\
&\qquad \left(\mathbb{E}\left(I_{\left\{(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk}-\beta_{jk}|^2)^{1/2} \geq \lambda/2\right\}}\right)\right)^{\frac{1}{2}} \left\|\psi_{jk}\right\|_{L_2}^2 \\
&\leq C \sum_{j=j_0}^{j_1} \left(\sum_{k=0}^{2^j-1} \mathbb{E}|\hat{\beta}_{jk} - \beta_{jk}|^4 \right)^{\frac{1}{2}} \\
&\qquad \left(P\left(\left(l_j^{-1} \sum_{k\in\mathscr{B}_{jK}} |\hat{\beta}_{jk} - \beta_{jk}|^2 \right)^{1/2} \geq \lambda/2 \right) \right)^{\frac{1}{2}}.
\end{aligned}
$$

Using our Proposition 2 and the Proposition 4.1 from Tribouley and Viennet [20] we get

$$
\begin{aligned}
G_{31} &\leq C \sum_{j=j_0}^{j_1} \frac{2^j}{n^2} \left(n^{-\frac{\theta}{2}} l_j^{-\frac{1}{2}(2+\theta)} 2^{\frac{j}{2}(2+\theta)} (\log n)^{1+\theta} \right)^{\frac{1}{2}} \\
&\leq C \frac{2^{j_1}}{n^{2+\frac{\theta}{4}}} \left(\frac{2^{j_1}}{l_{j_1}}\right)^{\frac{1}{4}(2+\theta)} (\log n)^{\frac{\theta+1}{2}} \\
&\leq C n^{-\frac{2s}{1+2s}}. \qquad (11)
\end{aligned}
$$

- The upper bound for G_{32}. Let j_3 an integer belonging to $[j_0, j_1]$ and M_1 such that $\|f\|_{\mathbf{H}^s} \leq M_1$. Since $\left(\sum_{j\geq\tau-1} 2^{2js} \left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 \right) \right)^{1/2} < \infty$. We have, for all $j \in [j_3, j_1]$

$$\left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2\right)^{1/2} \leq M_1 2^{-js} \leq M_1 2^{-j_3 s}. \tag{12}$$

Recall that $\lambda/2 = 2K_2/n^{1/2}$ and with choice of $2^{j_3} = \left(\frac{M_1}{2K_2}\right)^{\frac{2}{1+2s}} n^{\frac{1}{1+2s}}$. We will prove that by absurd. Assume that for all $j \in [j_3, j_1]$ we have $b_{jK} > \lambda/2$ then

$$\left(l_j^{-1} \sum_{k \in \mathscr{B}_{jK}} |\beta_{jk}|^2\right)^{\frac{1}{2}} > \lambda/2$$

$$l_j^{-1} \sum_{k \in \mathscr{B}_{jK}} |\beta_{jk}|^2 > (\lambda/2)^2$$

$$\sum_{k=0}^{2^j-1} |\beta_{jk}|^2 > (\lambda/2)^2 2^j$$

$$\left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2\right)^{1/2} > 2^{\frac{j}{2}} \lambda/2$$

$$\left(\sum_{k=0}^{2^j-1} |\beta_{jk}|^2\right)^{1/2} > 22^{\frac{j_3}{2}} \frac{K_2}{n^{\frac{1}{2}}} = M_1 2^{-j_3 s}$$

which contradicts (12). We conclude therefor, that the terms corresponding to $j \in [j_3, j_1]$ are zero. Thus

$$\begin{aligned} G_{32} &= \mathbb{E}\left\| \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} (\hat{\beta}_{jk} - \beta_{jk}) I_{\{\hat{b}_{jK} \geq \lambda\}} I_{\{b_{jK} > \lambda/2\}} \psi_{jk} \right\|_{L_2}^2 \\ &= \mathbb{E}\left\| \sum_{j=j_0}^{j_3} \sum_{k=0}^{2^j-1} (\hat{\beta}_{jk} - \beta_{jk}) I_{\{\hat{b}_{jK} \geq \lambda\}} I_{\{b_{jK} > \lambda/2\}} \psi_{jk} \right\|_{L_2}^2 . \end{aligned}$$

Using the Proposition 1, we get

$$
\begin{aligned}
G_{32} &= \mathbb{E}\left\| \sum_{j=j_0}^{j_3}\sum_{k=0}^{2^j-1}(\hat{\beta}_{jk}-\beta_{jk})I_{\{\hat{b}_{jK}\geq\lambda\}}I_{\{b_{jK}>\lambda/2\}}\psi_{jk}\right\|_{L_2}^2 \\
&\leq \mathbb{E}\left\| \sum_{j=j_0}^{j_3}\sum_{k=0}^{2^j-1}(\hat{\beta}_{jk}-\beta_{jk})\psi_{jk}\right\|_{L_2}^2 \\
&\leq C\sum_{j=j_0}^{j_3}\sum_{k=0}^{2^j-1}\mathbb{E}|\hat{\beta}_{jk}-\beta_{jk}|^2 \\
&\leq C\frac{2^{j_3}}{n} \leq Cn^{-\frac{2s}{1+2s}}.
\end{aligned} \tag{13}
$$

Combining the previous inequalities (4), (5), (8), (9), (11), and (13) in (3), we obtain the desired upper bounds. ∎

Proof of Theorem 2 The scheme of this proof is quite classical. We take inspiration from the proof of Theorem 2.2 in Tribouley and Viennet [20]. The proof is based on a results of Talagrand [19], the Berbee's lemma [1] with the fact that the β-mixing coefficient are less than ϕ-mixing coefficient and mainly on the following observation

Remark 1 The study of the quantity $\sum_{k\in\mathscr{B}_{jK}}|\hat{\beta}_{jk}-\beta_{jk}|^2$ is related to the study of supremum of a centered empirical process ν_n, where $\nu_n(\psi_{jk}) = \hat{\beta}_{jk}-\beta_{jk}$, over the set $F_2 = \left\{h = \sum_{k\in\mathscr{B}_{jK}} a_k\psi_{jk} / \sum_{k\in\mathscr{B}_{jK}}|a_k|^2 \leq 1\right\}$. By duality arguments we have

$$
\begin{aligned}
\sum_{k\in\mathscr{B}_{jK}}|\hat{\beta}_{jk}-\beta_{jk}|^2 &= \sup_{a_k,\sum_{k\in\mathscr{B}_{jK}}|a_k|^2\leq 1}\left|\sum_{k\in\mathscr{B}_{jK}}a_k(\hat{\beta}_{jk}-\beta_{jk})\right|^2 \\
&= \sup_{h\in F_2}|\nu_n(h)|^2.
\end{aligned}
$$

Lemma 1 (Berbee [1]) *Let $(X_i)_{i>0}$ be a sequence of random variables taking their values in a Polish space χ. Then, there exists a sequence $(X_i^*)_{i>0}$ of independent random variables such that for any positive integer i, we have*

$$
\mathbb{P}(X_i \neq X_i^*) < \beta(\sigma(X_j; j < i), \sigma(X_i)).
$$

Return now to the proof of Theorem 2. For the sake of simplicity, we assume in the following that $X_i = 0$ and $h(X_i) = 0$ if $i > n$, and that $E_P(h) = 0$. Let $q = [\sqrt{n}]$ be a fixed integer, where [.] denotes the integer part. According to Berbee's lemma, there exists a sequence of independent random variables $(X_i^*)_{i\geq 0}$ such that for any positive integer k, $Y_k = (X_{qk+1}, \ldots, X_{q(k+1)})$ and $Y_k^* = (X_{qk+1}^*, \ldots, X_{q(k+1)}^*)$ have the same distribution, and

$$\mathbb{P}(Y_k \neq Y_k^*) \leq \beta(q) \leq \phi(q).$$

Then, the random variables $(Y_{2k}^*)_{k\geq 0}$ are independent, and even $(Y_{2k+1}^*)_{k\geq 0}$.

Let ν_n^* be the empirical process associated with the random variables $(X_i^*)_{i\geq 0}$. The centered empirical process $\nu_n(h)$ (that is $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i} - P_{X_1}$) is then decomposed as

$$\nu_n(h) = \nu_n(h) - \nu_n^*(h) + \nu_n^*(h).$$

Since for any $\lambda_1 > 0$ and $\lambda_2 > 0$ we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{h\in F_2} |\nu_n(h)| \geq \lambda_1 + K_2 l_j^{\frac{1}{2}} \frac{1}{\sqrt{n}} + \lambda_2\right) \\
&\leq \mathbb{P}\left(\sup_{h\in F_2} |\nu_n(h) - \nu_n^*(h)| \geq \lambda_2\right) \\
&\quad + \mathbb{P}\left(\sup_{h\in F_2} |\nu_n^*(h)| \geq \lambda_1 + K_2 l_j^{\frac{1}{2}} \frac{1}{\sqrt{n}}\right). \qquad (14)
\end{aligned}$$

We just have to control these two terms to get the announced result.

For the first term in the right hand of (14), it is easy to have

$$|\nu_n(h) - \nu_n^*(h)| = \frac{1}{n}\left|\sum_{i=1}^n \delta_{X_i}(h) - \delta_{X_i^*}(h)\right| \leq \frac{2}{n}\|h\|_\infty \sum_{i=1}^n I_{X_i \neq X_i^*}.$$

We obtain then

$$\mathbb{E}|\nu_n(h) - \nu_n^*(h)| \leq \frac{2}{n}\,\|h\|_\infty\, n\,\phi(q).$$

Since $h \in F_2 = \left\{\sum_{k\in\mathscr{B}_{jK}} a_k\psi_{jk} / \sum_{k\in\mathscr{B}_{jK}} |a_k|^2 \leq 1\right\}$ and using Hölder's inequality, we have

$$|h(x)| \leq \left(\sum_{k\in\mathscr{B}_{jK}} |a_k|^2\right)^{\frac{1}{2}} \left(\sum_{k\in\mathscr{B}_{jK}} \left|\psi_{jk}(x)\right|^2\right)^{\frac{1}{2}} \leq \left(\sum_{k=0}^{2^j-1} \left|\psi_{jk}(x)\right|^2\right)^{\frac{1}{2}}, x \in [0,\ 1].$$

According to Lemma 6 in Birgé and Massart [2], $\left\| \sum_{k=0}^{2^j-1} \psi_{jk}^2 \right\|_\infty \leq C\, 2^j$, we get that

$$\|h\|_\infty \leq C\, 2^{\frac{j}{2}}.$$

Therefore

$$\mathbb{E}|\nu_n(h) - \nu_n^*(h)| \leq 2\|h\|_\infty \phi(q) \;\leq\; 2\, C\, 2^{\frac{j}{2}}\, \phi(q).$$

By Markov inequality, we deduce that for any $\lambda_2 > 0$

$$\mathbb{P}\left(\sup_{h \in F_2} |\nu_n(h) - \nu_n^*(h)| \geq \lambda_2 \right) \leq \frac{2C}{\lambda_2} 2^{\frac{j}{2}} \phi(q). \tag{15}$$

For the second term in the right hand of (14), we use the following theorem.

Theorem 3 (Talagrand [19]) *Let $X_1, \ldots, X_n$, be n independent identically distributed random variables, and F_2 a family of functions that are uniformly bounded by some constant M_1. Let V and H be defined by*

$$\frac{1}{n}\mathbb{E}\left(\sup_{h \in \mathrm{F}_2} \left| \sum_{i=0}^n h(X_i) \right| \right) \leq H, \qquad V = \mathbb{E}\left(\sup_{h \in \mathrm{F}_2} \sum_{i=0}^n h^2(X_i) \right).$$

Then, there exists a universal constant K_1 such that for any $\lambda_1 > 0$,

$$\mathbb{P}\left(\sup_{h \in \mathrm{F}_2} |\nu_n(h)| \geq \lambda_1 + H \right) \leq \exp\left[-K_1 n \left(\frac{\lambda_1^2}{V} \wedge \frac{\lambda_1}{M_1} \right) \right].$$

Moreover, according to Corollary 3.4 in Talagrand [18], the following bound holds

$$V \leq v + 8M_1\tilde{H},$$

where v and $\tilde{H}$ are defined by

$$\sup_{h \in \mathrm{F}_2} \mathrm{Var}(h(X)) \leq v, \qquad \frac{1}{n}\mathbb{E}\left(\sup_{h \in \mathrm{F}_2} \sum_{i=0}^n \varepsilon_i h(X_i) \right) \leq \tilde{H},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are n independent Rademacher variables.

We derive then the following inequality: there exists a positive constant k_1 such that for any $\lambda_1 > 0$,

$$\mathbb{P}\left(\sup_{h\in \mathrm{F}_2} |\nu_n(h)| \geq \lambda_1 + H\right) \leq \exp\left[-k_1 n\left(\frac{\lambda_1^2}{v} \wedge \frac{\lambda_1}{M_1} \wedge \frac{\lambda_1^2}{8M_1\tilde{H}}\right)\right].$$

Let $p = p(n)$ be the greatest integer such that $n = qp + r$ with $0 \leq r < q$, then

$$\nu_n^*(h) = \frac{p}{n}\left[\frac{1}{p}\sum_{l=0}^{p}\left(\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*) + \sum_{i=q(2l+1)+1}^{q(2l+2)} h(X_i^*)\right)\right].$$

The control of the second term in the right hand of (14) is made in two steps, considering odd terms and even ones. They are both treated in the same way, so we only detail the even part. Since the variables $\left(\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)\right)_{0\leq l\leq p}$ are independent by construction, we are allowed to apply the Talagrand's inequality to $\nu_p^* = \frac{1}{p}\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)$, with adequate choices of λ_1, H, $\tilde{H}$, M_1 and v. Let F_2 be the family of functions defined previously. We have to determine the quantities M_1, $\tilde{H}$, H and v. Since

$$\left\|\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)\right\|_\infty \leq q\,\|h\|_\infty \leq C\,q\,2^{\frac{j}{2}},$$

we put

$$M_1 = C\,q 2^{\frac{j}{2}}. \tag{16}$$

Let us now determine the quantity H. Applying Hölder's inequality and Jensen's inequality, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{h\in F_2}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*)\right|\right]\\
&= \mathbb{E}\left[\sup_{h\in F_2}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\sum_{k\in\mathscr{B}_{jK}} a_k\psi_{jk}(X_i^*)\right|\right]\\
&\leq \mathbb{E}\left[\sup_{h\in F_2}\left(\sum_{k\in\mathscr{B}_{jK}}|a_k|^2\right)^{\frac{1}{2}} \times \left(\sum_{k\in\mathscr{B}_{jK}}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right|^2\right)^{\frac{1}{2}}\right]
\end{aligned}$$

$$\leq \mathbb{E}\left[\sum_{k\in\mathscr{B}_{jK}}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right|^2\right]^{\frac{1}{2}}$$

$$\leq \left[E\sum_{k\in\mathscr{B}_{jK}}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right|^2\right]^{\frac{1}{2}}.$$

In order to bound this last term we have, under the summability condition on the mixing coefficients $B<\infty$ and the fact that the variables $\left(\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right)_{0\leq l\leq p}$ are independent and satisfy $\mathbb{E}\left(\psi_{jk}(X_i^*)\right)=0$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{k\in\mathscr{B}_{jK}}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right|^2\right] &= \sum_{k\in\mathscr{B}_{jK}}\mathbb{E}\left[\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}\psi_{jk}(X_i^*)\right|^2\right]\\
&\leq \sum_{k\in\mathscr{B}_{jK}} p\,4\,\|\psi\|_{L_1}\|\psi\|_\infty\|f\|_\infty q\sum_{i=2ql+1}^{q(2l+1)}\phi(i)\\
&\leq 4l_j\,\|\psi\|_{L_1}\|\psi\|_\infty\|f\|_\infty p\,q\sum_{i\geq 0}\phi(i)\\
&\leq 4B\,\|\psi\|_{L_1}\|\psi\|_\infty\,\|f\|_\infty\,l_j\,n.
\end{aligned}$$

We deduce then

$$\frac{1}{p}\mathbb{E}\left[\sup_{h\in F_2}\left|\sum_{l=0}^{p}\sum_{i=2ql+1}^{q(2l+1)}h(X_i^*)\right|\right]\leq K_2\frac{\sqrt{n}}{p}l_j^{\frac{1}{2}}.$$

We take, now

$$H=K_2\sqrt{\frac{q}{p}}\,l_j^{\frac{1}{2}} \tag{17}$$

with

$$K_2=\left(4B\,\|\psi\|_{L_1}\|\psi\|_\infty\,\|f\|_\infty\right)^{\frac{1}{2}}.$$

We notice that

$$\frac{1}{n}\mathbb{E}\left(\sup_{h\in F_2}\sum_{i=1}^{n}\epsilon_i h(X_i)\right)\leq K_2 l_j^{\frac{1}{2}}\sqrt{\frac{q}{p}},$$

and we choose

$$\tilde{H} = H. \tag{18}$$

Let us now determine the quantity v such that $\sup_{h \in F_2} \mathrm{Var}(h(X)) \le v$.

Let $h \in \left\{ \sum_{k \in \mathscr{B}_{jK}} a_k \psi_{jk} / \sum_{k \in \mathscr{B}_{jK}} |a_k|^2 \le 1 \right\}$, we get

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=2ql+1}^{q(2l+1)} h(X_i^*) \right) &\le q K_2^2 \int h^2 f \\
&\le q K_2^2 \int \left(\sum_{k \in \mathscr{B}_{jK}} a_k \psi_{jk} \right)^2 f \\
&\le q K_2^2 \int \left(\left(\sum_{k \in \mathscr{B}_{jK}} |a_k|^2 \right)^{\frac{1}{2}} \left(\sum_{k \in \mathscr{B}_{jK}} \psi_{jk}^2 \right)^{\frac{1}{2}} \right)^2 f \\
&\le q K_2^2 \int \sum_{k \in \mathscr{B}_{jK}} \psi_{jk}^2 f \\
&\le K_2^2 \|f\|_\infty q l_j.
\end{aligned}$$

Thus we set

$$v = K_2^2 \|f\|_\infty q l_j. \tag{19}$$

Finally, applying Talagrand's inequality with H, $\tilde{H}$, M_1 and v defined in (17), (18), (16), and (19) , we obtain

$$\begin{aligned}
&\mathbb{P}\left(\sup_{h \in F_2} \left| \frac{1}{p} \sum_{l=0}^{p} \sum_{i=2ql+1}^{q(2l+1)} h(X_i^*) \right| \ge q\lambda_1 + K_2 l_j^{\frac{1}{2}} \sqrt{\frac{q}{p}} \right) \\
&\quad \le \exp\left(-k_1 \mathbb{P}\left(\frac{q^2 \lambda_1^2}{4K_2^2 \|f\|_\infty q l_j} \wedge \frac{q\lambda_1}{C\, q 2^{\frac{j}{2}}} \wedge \frac{q^2 \lambda_1^2}{8C\, q 2^{\frac{j}{2}} K_2 l_j^{\frac{1}{2}} \sqrt{\frac{q}{p}}} \right) \right)
\end{aligned}$$

$$\leq \exp\left(-k_1 n\left(\frac{\lambda_1^2 l_j^{-1}}{4K_2^2\|f\|_\infty} \wedge \frac{\lambda_1}{C\,q2^{\frac{j}{2}}} \wedge \frac{\lambda_1^2\sqrt{n}}{ql_j^{\frac{1}{2}}2^{\frac{j}{2}}}\frac{1}{8CK_2}\right)\right)$$

$$\leq \exp\left(-K_1 n\left(\lambda_1^2 l_j^{-1} \wedge \frac{\lambda_1}{q2^{\frac{j}{2}}} \wedge \frac{\lambda_1^2\sqrt{n}}{ql_j^{\frac{1}{2}}2^{\frac{j}{2}}}\right)\right), \tag{20}$$

where K_1 is a positive constant depending on $\|f\|_\infty$, K_2, C and k_1. Similarly, we find for the odd part

$$\mathbb{P}\left(\sup_{h\in F_2}\left|\frac{1}{p}\sum_{l=0}^{p}\sum_{i=q(2l+1)+1}^{q(2l+2)} h(X_i^*)\right| \geq q\lambda_1 + K_2 l_j^{\frac{1}{2}}\sqrt{\frac{q}{p}}\right)$$

$$\leq \exp\left(-K_1 n\left(\lambda_1^2 l_j^{-1} \wedge \frac{\lambda_1}{q2^{\frac{j}{2}}} \wedge \frac{\lambda_1^2\sqrt{n}}{ql_j^{\frac{1}{2}}2^{\frac{j}{2}}}\right)\right). \tag{21}$$

Regrouping (15), (21), and (20), the proof of Theorem 2 is complete. ■

References

1. Berbee, H.C.P.: Random walks with stationary increments and renewal theory. Cent. Math. Tracts, Amsterdam (1979)
2. Birgé, L., Massart, P.: Minimum contrast estimators on sieves: exponential bounds and rates of convergence. Bernoulli **4**(3), 329–375 (1998)
3. Cai, T.: Minimax wavelet estimation via block thresholding. Technical Report, pp. 96–41, Department of Statistics, Purdue University (1996)
4. Cai, T.: On adaptivity of blockshrink wavelet estimator over Besov spaces. Technical Report, pp. 97–05, Department of Statistics, Purdue University (1997)
5. Cai, T.: Adaptive wavelet estimation: a block thresholding and oracle inequality approach. Ann. Stat. **27**, 898–924 (1999)
6. Cai, T.: On block thresholding in wavelet regression: adaptivity, block size and threshold level. Stat. Sin. **12**, 1241–1273 (2002)
7. Chesneau, C.: Quelques contributions à l'estimation fonctionnelle par méthodes d'ondelettes. Ph.D. Thesis, Université Pierre et Marie Curie, Paris VI (2006)
8. Chesneau, C.: Wavelet estimation via block thresholding: a minimax study under Lp-risk. Stat. Sin. **18**, 1007–1024 (2008)
9. Efroimovich, S.Y.: Nonparametric estimation of a density of unknown smoothness. Theory Probab. Appl. **30**, 557–561 (1985)
10. Hall, P., Kerkyacharian, G., Picard, D.: Block thresholding rules for curve estimation using kernel and wavelet methods. Ann. Stat. **26**, 942–962 (1998)
11. Hall, P., Kerkyacharian, G., Picard, D.: On the minimax optimality of block thresholded wavelet estimators. Stat. Sin. **9(1)**, 33–49 (1999)
12. Ibragimov, I.A.: Some limit theorems for stationary processus. Theor. Probab. Appl. **7**, 349–382 (1962)

13. Kerkyacharian, G., Picard, D., Tribouley, K.: Density estimation in Besov spaces. Bernoulli **2**, 229–247 (1996)
14. Mallat, S.: Une exploration des signaux en ondelettes. Ecole Polytechnique Ellipses Edition (2000)
15. Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. Academic, New York (2008)
16. Meyer, Y.: Ondelettes et Opérateurs. Hermann, Paris (1990)
17. Rio, E.: Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiques & Applications, vol. 31. Springer, Berlin (2000)
18. Talagrand, M.: Sharper bounds for gaussian and empirical processes. Ann. Probab. **22**, 28–76 (1994)
19. Talagrand, M.: New concentration inequalities in product spaces. Technical Report, University Paris VI, Ohio State University (1995)
20. Tribouley, K., Viennet, G.: Lp-adaptive density estimation in a β-mixing framework. Ann. Inst. Henri Poincarée **34**, 179–208 (1998)

Approximation of Strictly Stationary Banach-Valued Random Sequence by Fourier Integral

Tawfik Benchikh

Abstract This paper is devoted to the approximation of a second-order E-valued strictly stationary random sequence by the Fourier transform of a L_E^2-valued random measure, where E is a complex separable Banach space. For this purpose, we use the spectral representation of a second order E-valued stationary random function and we introduce a bijective linear operator on L_E^2 which preserves the norm in the form of a "shift operator" associated with a L_E^2-valued strictly stationary sequence.

1 Introduction

In recent decades, multidimensional statistical methods have known very important developments in both theoretical aspects (particularly in operatorial statistics) and practical applications, e.g. enabling functional data treatments. Indeed, the interpretation of a process as elements with values in functional space has become a useful tool in the analysis of functional data. For example, to a real continuous process X we associate the Banach-valued process $X(.)$ called a window process, which describes the evolution of X taking into account a memory $\rho > 0$. The natural state space for $X(.)$ is the Banach space of continuous functions on $[0, \rho]$. A recent reference for Banach space-valued processes is Bosq [5].

On the other hand, the frequency domain's field (and its important tool the Fourier transform) permits the analysis of the stationary random function, which is very useful in the descriptive study, interpolation, estimation and prediction problems of the such processes. The spectral point of view is particularly advantageous in the analysis of multivariate stationary processes and in the analysis of very large data sets, for which numerical calculations can be performed rapidly using the fast Fourier transform or the inverse transform.

There exists a unique $\mathbb{C}^p$-valued random measure Z, called p-random measure associated to $(X_g)_{g\in G}$; defined on the Borel σ-field $\mathscr{B}_{\hat{G}}$ of dual of G, and taking

T. Benchikh (✉)
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, BP 89, Sidi Bel Abbés 22000, Algeria
e-mail: benchikh@yahoo.fr

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*, Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_3

values in $L^2_{\mathbb{C}^p}$, such that $X_g = \int_{\hat{G}} (., g)_{\hat{G}G} I_{\mathbb{C}^p} dZ$, for any $g \in G$, where $(\gamma, g)_{\hat{G}G}$ denote the value of $\gamma \in \hat{G}$ at $g \in G$ (see Azencott and Dacunha-Castelle [1], Boudou and Dauxois [7]). Conversely, if Z is a p-random measure defined on $\mathscr{B}_{\hat{G}}$ then $(\int_{\hat{G}} (., g)_{\hat{G}G} I_{\mathbb{C}^p} dZ)_{g\in G}$ is a p-dimensional stationary continuous random function.

For $G = \mathbb{Z}$, we have $\hat{G} =]-\pi, \pi]$ and $(Z(\lambda))_{\lambda \in]-\pi,\pi]}$ is a p-dimensional spectral process with orthogonal increments, i.e., the spectral representation of stationary process $(X_n)_{n\in\mathbb{Z}}$ essentially decompose (X_n) into a sum of sinusoidal components with uncorrelated random coefficients. For example, a complex-valued stationary process has a decomposition as:

$$X_n = \sum_k A_k \cos(\lambda_k n + \phi)$$

where A_k are random amplitude, and ϕ random phases independent of the A_k.

The spectral tools can be used for estimation and predictive process models for spatial and spatiotemporal data (see Yalgom [15]), or for the interpolation problem of stationary sequences (see Boudou [6]), or else to study processes $(T_n)_{n\in\mathbb{Z}}$ which may be written multiplicatively with two independent components, $(T_n)_{n\in\mathbb{Z}} = (X_n Y_n)_{n\in\mathbb{Z}}$, where $(X_n)_{n\in\mathbb{Z}}$ and $(Y_n)_{n\in\mathbb{Z}}$ are independent stationary multivariate sequences (see Boudou and Romain [9], Tensor and convolution products of random measures). The descriptive study (i.e. Principal component analysis in the frequency domain, see Brillinger [4]) and the simulation of stationary (spatial) processes are other topics.

Then, it seems useful to study the spectral tools associated to a Banach space-valued stationary random function.

The spectral theory of operator-valued stationary continuous random function on abelian locally compact groups G (with dual space $\hat{G}$ admitting a countable basis), is well developed (see Chobanyan and Weron [10]) and is also intensively studied by various authors. Indeed, given a continuous stationary random function $(K_g)_{g\in G}$ of elements of $\mathscr{L}(L^2, E)$, there exists a unique operator-valued random measure Z defined on the Borel σ-field $\mathscr{B}_{\hat{G}}$, and taking values in $\mathscr{L}(L^2, E)$, such that $(K_g)_{g\in G}$ is Fourier transform of Z. When E is a Hilbert space, such representations of random functions were considered in detail by Payen [13], Masani [12], Kallianpur and Mandrekar [11], Yalgom [15], among others.

In our paper [3], the relationship between a L^2_E-valued random measures and continuous second-order E-valued stationary random function on abelian locally compact groups G of elements L^2_E has been proved under the previous assumption. Our approach is based on the technique used in Azencott and Dacunha-Castelle [1] and after by Boudou and Romain [8].

This results is used (Sect. 4) to approximate a second-order E-valued strictly stationary random sequence by the Fourier transform of a random measure Z taking values in L^2_E. For this purpose, we will introduce the transformation that is present in the form of a bijective isometry associated with a L^2_E-valued strictly stationary sequence.

2 Notations and Preliminaries

In the whole paper, E is a separable Banach space with topological dual E', and H will be a separable complex Hilbert space equipped with scalar product $< .,. >$. The topological dual of H is denoted by H' and the antilinear isomorphism $h \in H \mapsto < ., h > \in H'$ by $\mathscr{I}$. We denote by $\mathscr{K}(H, E)$ the Banach space of compacts linear operators from H to E. If $K \in \mathscr{L}(H, E)$, then ${}^tK$ stands for its transposed operator.

For an operator K of $\mathscr{L}(H, E)$, we put ${}^qK = \mathscr{I}^{-1} \circ {}^tK$. The operator ${}^qK$, which will be called the *quasi-transpose operator* of K, is an antilinear one and, for all $(e', h) \in E' \times H$, we have: $< h, {}^qKe' >= (e', Kh)_{E'E}$, where $(., .)_{E'E}$ denotes the duality between E and E'.

For any two Banach spaces E and F, $\mathscr{L}(E, F)$ denotes the Banach space of all continuous linear operators from E to F, $a : E \to F$, endowed with the norm $\|a\| = \sup_{\|e\| \leq 1} \|a(e)\|$.

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a fixed probability space such that the complex Banach space $L^2_E = L^2_E(\Omega, \mathscr{A}, \mathbb{P})$ is separable. When $E = \mathbb{C}$, we will simply write L^2.

When X is an element of L^2_E, the mapping $\tilde{X} : y \in L^2 \mapsto \mathbb{E}(\underline{yX}) \in E$, is a compact operator. Its quasi-transpose $\tilde{X}$ is defined by ${}^q\tilde{X} : e' \in E' \mapsto \overline{e' \circ X} \in L^2$. Hence, we can verify that

Proposition 1 *For all $(\varphi, X) \in \mathscr{L}(E, F) \times L^2_E$, $\varphi \circ X \in L^2_F$ and that $\widetilde{\varphi \circ X} = \varphi \circ \tilde{X}$.*

Note that if E is an infinite-dimensional space, then there exists an operator $K \in \mathscr{K}(L^2, E)$ which is not associated with random element X with values in Banach space E, i.e., there may not exist a random element $X \in L^2_E$ such that $K = \tilde{X}$ (see Chobanyan et Weron [10]). A sufficient condition for $K \in \mathscr{K}(L^2, E)$ to be associated with a second-order E-valued random element is given by the following proposition.

Proposition 2 (cf. [10]) *Let $K \in \mathscr{K}(L^2, E)$ be such that it* factors through a Hilbert–Schmidt operator, *i.e., admit a factorization $K = \varphi \circ \psi$, where H_1 is a complex Hilbert space, $\psi \in \mathscr{L}(L^2, H_1)$ is a Hilbert–Schmidt operator, and $\varphi \in \mathscr{L}(H_1, E)$. Then there is an random element $X \in L^2_E$ such that $K = \tilde{X}$.*

For example, the operator $\sum_{i=1}^{p} y_i \otimes e_i$, where $(y_i, e_i) \in L^2 \times E$ is factorized through an operator of Hilbert–Schmidt.

In [2], we give a sufficient condition for the operator $K \in \mathscr{K}(H, E)$ to be factorized through a Hilbert–Schmidt operator.

Let G be an additive locally compact abelian (LCA) group with the dual $\hat{G}$ and let $(\gamma, g)_{\hat{G}G}$ denote the value of $\gamma \in \hat{G}$ at $g \in G$. The Borel σ-field of $\hat{G}$ is denoted by $\mathscr{B}_{\hat{G}}$. We suppose that $\hat{G}$ admits a countable basis.

A continuous compact operator-valued stationary random function $(K_g)_{g\in G}$ on G is a function $K : G \to \mathscr{K}(H, E)$ such that

- the mapping $g \in G \to (K_g \circ {}^q K_0 e')(e')$ is continuous for each $e' \in E'$, and
- for any pairs (g, g') of elements of $G \times G$, we have $K_g \circ {}^q K_{g'} = K_{g-g'} \circ {}^q K_0 = K(g - g')$.

It is well known (see Chobanyan and Weron [10]) that each continuous compact operator-valued stationary random function $(K_g)_{g\in G}$ has the representation $K_g = \int (., g)_{\hat{G},G} dZ$, for all $g \in G$, where Z is a vector measure on $\mathscr{B}_{\hat{G}}$ with values in $\mathscr{K}(H, E)$ such that $Z(A) \circ {}^q(Z(B)) = 0$, for any pairs (A, B) of disjoint elements of $\mathscr{B}_{\hat{G}}$. The measure Z is called the *operator-valued random measure associated with the* $(K_g)_{g\in G}$.

For the same properties of the integral with respect to an operator-valued random measure, we can see [3]. Hence, it can be shown that:

Corollary 1 *Two operator-valued random measures Z_1 and Z_2 are equal if $\int (, g)_{\hat{G},G} dZ_1 = \int (, g)_{\hat{G},G} dZ_2$, for all $g \in G$.*

3 Spectral Representation of Banach-valued Stationary Random Function

In this section we recall the spectral representation of L^2_E-valued stationary random processes (see Benchikh [3]). Adopting the notations of Sect. 2, we can introduce the following definition.

Definition 1 A Banach space-valued random measure Z is a mapping defined on $\mathscr{B}_{\hat{G}}$ and taking values in L^2_E such that

(i) for all pairs (A, B) of disjoint elements of $\mathscr{B}_{\hat{G}}$, we have $Z(A \cup B) = ZA + ZB$ and $\widetilde{ZA} \circ {}^q\widetilde{ZB} = 0$;
(ii) for each decreasing sequence $(A_n)_{n\in\mathbb{Z}}$ of elements of $\mathscr{B}_{\hat{G}}$ converging to $\varnothing$, we have: $\lim_{n\mapsto\infty} \|ZA_n\|_{L^2_E} = 0$.

It is easy to show that if Z is a Banach space-valued random measure, then the mapping $\tilde{Z} : A \in \mathscr{B} \mapsto \widetilde{ZA} \in \mathscr{K}(L^2, E)$ is an operator-valued random measure called the *operator-valued random measure associated with* Z.

Recall now the definition of the continuous second-order Banach-valued stationary random function (see Bosq [5]).

Definition 2 We say that a family $(X_g)_{g\in G}$ of elements of L^2_E is a stationary continuous random function if, for all $e' \in E'$, the family $(e' \circ X_g)_{g\in G}$ is stationary continuous random function and, for all pairs (e', f') of elements of E', the families $(e' \circ X_g)_{g\in G}$ and $(f' \circ X_g)_{g\in G}$ are stationarily correlated.

Thus, if $(X_g)_{g\in G}$ is a L^2_E-valued stationary continuous random function, then we have

$$< e' \circ X_g, f' \circ X_{g'} >=< {}^q\widetilde{X_{g'}}f', {}^q\widetilde{X_g}e' >= (e', \widetilde{X_g} \circ {}^q\widetilde{X_{g'}}f')_{E',E}$$

for all $(e',f') \in E'^2$, $(g,g') \in G^2$.

It can be said also that a family $(X_g)_{g\in G}$ of elements of L^2_E is a stationary continuous random function if and only if the family $(\widetilde{X_g})_{g\in G}$ is stationary.

We know that for a stationary family $(K_g)_{g\in G}$, there exists one and only one operator-valued random measure Z, such that $K_g = \int (.,g)_{\hat{G},G} dZ$, for all $g \in G$. However, we have seen that when E is an infinite-dimensional space, then there exist an operator $K \in \mathscr{L}(L^2, E)$ which is not associated with random element X with values in Banach space E. Then there may exist a stationary continuous random function $(X_g)_{g\in G}$ of elements of L^2_E for which it cannot be asserted the existence of a Banach space-valued random measure Z such that $X_g = \int (\cdot, g)_{\hat{G}G} dZ$ for all $g \in G$. Nevertheless, if there is one, it is unique (Corollary 1).

Such a condition is given by the Proposition 2. Consequently, we can give the following results.

Proposition 4 (cf. [3]) *If $(X_g)_{g\in G}$ is a stationary continuous random function of elements of L^2_E, such that $\widetilde{X_0}$ is factorized through a Hilbert–Schmidt operator, then there exists a L^2_E-valued random measure Z such that $X_g = \int (.,g)_{\hat{G}G} dZ$, for all $g \in G$.*

4 Approximation of a Second-order Strictly Stationary Banach-valued Random Sequence by Banach-valued Random Measure

The purpose of this section is the study of a particular form of stationarity, the strictly stationary random sequence. In the following we put $G = \mathbb{Z}$. Then the dual $\hat{G}$ is identified with $\Pi = [-\pi, \pi[$.

Consider the space $F = E^{\mathbb{Z}}$ equipped with the smallest σ-algebra $\mathscr{F}$ that makes measurable the coordinate mapping $T_n : (e_m)_{m\in\mathbb{Z}} \in F \mapsto e_n \in E$, $n \in \mathbb{Z}$.

On the space F there is the natural shift defined by $\Theta : (e_n)_{n\in\mathbb{Z}} \in F \mapsto (e_{n+1})_{n\in\mathbb{Z}} \in F$. Clearly, the mapping Θ is a measurable map and we have $\Theta^{-1}(\mathscr{F}) = \mathscr{F}$.

Any Banach-valued random sequence $(X_n)_{n\in\mathbb{Z}}$ of elements of L^2_E induce a measurable mapping

$$\check{X} : \omega \in \Omega \mapsto (X_n(\omega))_{n\in\mathbb{Z}} \in F$$

called the trajectory (or sample path) random variable associated to $(X_n)_{n\in\mathbb{Z}}$. Then, we can say that a random sequence $(X_n)_{n\in\mathbb{Z}}$ of elements of L^2_E is strictly stationary if $\Theta(\check{X}(\mathbb{P})) = \check{X}(\mathbb{P})$ (the law of associate trajectory r.v. $\check{X}$ is invariant by Θ).

Naturally when $E = \mathbb{C}$, we find the classical definition of a strictly stationary sequence.

4.1 Shift Operator Associated to a Strictly Stationary Banach-valued Random Sequence

Let $(X_n)_{n\in\mathbb{Z}}$ be a second-order strictly stationary Banach-valued random sequence. As the map $\check{X}$ is measurable from Ω into F, then, one can easily show the following result (see Benchikh et al. [2]):

Proposition 5

(*i*) *The mapping* $\mathscr{J} : T \in L^2_E(F, \mathscr{F}, \check{X}(\mathbb{P})) \mapsto T \circ \check{X} \in L^2_E(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$ *is a bijective linear which preserves the norm (isometry);*

(*ii*) *the mapping* $J : t \in L^2(F, \mathscr{F}, \check{X}(\mathbb{P})) \mapsto t \circ \check{X} \in L^2(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$ *is an isometry;*

(*iii*) *for all* $T \in L^2_E(F, \mathscr{F}, \check{X}(\mathbb{P}))$, *we have:* $\widetilde{\mathscr{J}T} = \tilde{T} \circ J^{-1}$ *and, for all* $W \in L^2(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$, *we have:* $\widetilde{\mathscr{J}^{-1}W} = \tilde{W} \circ J$.

Similarly, given that Θ is a measurable map from F into itself which checks $\Theta^{-1}(\mathscr{F}) = \mathscr{F}$ and $\Theta(\check{X}(\mathbb{P})) = \check{X}(\mathbb{P})$, we obtain the following satisfied proposition.

Proposition 6

(*i*) *The mapping* $\mathscr{V} : T \in L^2_E(F, \mathscr{F}, \check{X}(\mathbb{P})) \mapsto T \circ \Theta \in L^2_E(F, \mathscr{F}, \check{X}(\mathbb{P}))$ *is a bijective linear which preserve the norm (isometry);*

(*ii*) *the mapping* $V : t \in L^2(F, \mathscr{F}, \check{X}(\mathbb{P})) \mapsto t \circ \Theta \in L^2(F, \mathscr{F}, \check{X}(\mathbb{P}))$ *is a unitary operator.*

(*iii*) *for all* $T \in L^2_E(F, \mathscr{F}, \check{X}(\mathbb{P}))$, *we have:* $\widetilde{\mathscr{V} \circ T} = \tilde{T} \circ V^{-1}$ *and* $\widetilde{\mathscr{V}^{-1}T} = \tilde{T} \circ V$.

Consequently, we can deduce that the mapping $\mathscr{U} = \mathscr{J} \circ \mathscr{V} \circ \mathscr{J}^{-1}$ is a bijective linear operator of $L^2_E(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$ which preserves the norm, and $U = J \circ V \circ J^{-1}$ is a unitary operator of $L^2(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$.

The following property enables the establishment of a relationship between the operators $\mathscr{U}$ and U.

Proposition 7 *For all* $Y \in L^2_E(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$, *we have:*

$$\widetilde{\mathscr{U}Y} = \tilde{Y} \circ U^{-1} and \widetilde{\mathscr{U}^{-1}(Y)} = \tilde{Y} \circ U.$$

By convention, we set $K^0 = I$ the identity operator, and, when n is a negative integer, we write $K^n = (K^{-1})^{-n}$. Then, from the last property, we can deduce the following result

Proposition 8 *Given $Y \in L^2_E(\Omega, \check{X}^{-1}(\mathscr{F}), \mathbb{P})$, we can affirm that $\widetilde{\mathscr{U}^n Y} = \tilde{Y} \circ U^{-n}$, for all $n \in \mathbb{Z}$, and that $(\mathscr{U}^n Y)_{n\in\mathbb{Z}}$ is a stationary sequence.*

Consequently, the random sequence $(X_n)_{n\in\mathbb{Z}} = (\mathscr{U}^n X_0)_{n\in\mathbb{Z}}$ is stationary and $\widetilde{X_n} = \widetilde{X_0} \circ U^{-n}$, for all $n \in \mathbb{Z}$.

Remark 1 The operator $\mathscr{U}$ is an operator of shift. Indeed, for all (S, n) of $\mathscr{L}(E)\times\mathbb{Z}$, we have: $\mathscr{U}(S \circ X_n) = S \circ X_{n+1}$.

4.2 Approximation of a Second-order Strictly Stationary Banach-valued Random Sequence

Let us suppose that E has a Schauder basis denoted $\{x_k; k \in \mathbb{N}\}$, i.e., for all e of E, there exists a unique sequence of scalars $(\alpha_n)_{n\in\mathbb{N}} \in \mathbb{C}$ such that $e = \lim\limits_{p\to\infty} \sum_{k=0}^{p} \alpha_k x_k$ (convergence holds in the strong topology of E), and let $\{f_k; k \in \mathbb{N}\}$ the sequence of coefficient functionals associated with the basis $\{x_k; k \in \mathbb{N}\}$. If we set $S_p = \sum_{k=0}^{p} f_k \otimes x_k$, the sequence of partial sum operators associated to the basis $\{x_k; k \in \mathbb{N}\}$, then we have, for all e of E, $e = \lim\limits_{p} S_p(e)$ and $\|e\| \le \sup\{\|S_p e\|; p \in \mathbb{N}\} \le C\|e\|$ (see. Singer [14]). So, we can prove the following result:

Proposition 9 *For all X of L^2_E, $\lim\limits_{p} S_p \circ X = X$.*

Proof Indeed, as $X(\omega) = \lim\limits_{p} S_p(X(\omega))$, for all $\omega \in \Omega$, then we have:

$$\lim_{p} \|X(\omega) - S_p(X(\omega))\|^2 = 0.$$

Moreover as $\|X(\omega) - S_p(X(\omega)\|^2 \le (1 + C)^2 \|X(\omega)\|^2$, the dominated convergence theorem enables us to write that:

$$\begin{aligned}\lim_{p} \|X - S_p \circ X\|^2 &= \lim_{p} \int \|X(\omega) - S_p(X(\omega))\|^2 d\,\mathbb{P}(\omega)\\ &= \int \lim_{p} \|X(\omega) - S_p(X(\omega))\|^2 d\,\mathbb{P}(\omega) = \int 0\, d\,\mathbb{P} = 0\,,\end{aligned}$$

from which the result. □

According to this, we can approximate a L^2_E-valued strictly stationary sequence by a Fourier integral of a L^2_E-valued random measure.

Proposition 10 *Let $(X_n)_{n\in\mathbb{Z}}$ a strictly stationary sequence of L^2_E-valued random variables. Then, for each $\epsilon \in \mathbb{R}^*_+$, we may associate a b.r.m. Z_ϵ such that $\|X_n - \int e^{i.n} dZ_\epsilon\| \le \epsilon$, for all n of $\mathbb{Z}$.*

Proof Let us consider a strictly stationary sequence $(X_n)_{n\in\mathbb{Z}}$ of elements of L^2_E and let ϵ be a strictly positive real number. As the sequence $(S_p \circ X_0)_{p\in\mathbb{N}}$ converge to X_0, thus there exists a number p such that $\|X_0 - S_p \circ X_0\| \leq \epsilon$.

The Propositions 1 and 8 attest that the random sequence $(S_p \circ X_n)_{n\in\mathbb{Z}}$ is stationary and that the operator $\widetilde{S_p \circ X_0} = S_p \circ \widetilde{X_0} = \sum_{k=0}^{p} ({}^q\widetilde{X_0} f_k) \otimes x_k$ is factorized through an operator of Hilbert–Schmidt. According to the Proposition 4, there exists a Banach space-valued random measure Z_ϵ such that $S_p \circ X_n = \int e^{i.n} dZ_\epsilon$, for all $n \in \mathbb{Z}$.

Then, for all n of $\mathbb{Z}$, we have:

$$\begin{aligned}\|X_n - \textstyle\int e^{i.n} dZ_\epsilon\| &= \|X_n - S_p \circ X_n\| \\ &= \|\mathscr{U}^n(X_0) - \mathscr{U}^n(S_p \circ X_0)\| \\ &= \|X_0 - S_p \circ X_0\| \leq \epsilon\,.\end{aligned}$$

□

References

1. R. Azencott, D. Dacunha-Castelle: Séries d'observations irrégulières. Modélisation et prévision. Techniques stochastiques, Masson, Paris (1984)
2. T. Benchikh, A. Boudou, Y. Romain: Mesures aléatoires banachiques. Publi. Labo. Stat. Proba. 12-2006. Université Toulouse III - Paul Sabatier, Toulouse (2006)
3. T. Benchikh: Spectral representation of Banach-valued stationary random function on locally compact Abelian group. Georgian Math. J. **21**(2), 139–145 (2014)
4. D. Brillinger: Time Series Analysis and Theory. Society for Industrial Applied Mathematics, Philadelphia (2001)
5. D. Bosq: Linear Processes in Function Spaces: Theory and Mappings. Lecture Notes in Statistics, vol. 149. Springer, New York (2000)
6. A. Boudou: Interpolation de processus stationnaire. C.R.A.Sc. Paris Ser. I **336**(12), 1021–1024 (2003)
7. A. Boudou, J. Dauxois: Principal component analysis for a stationary random function defined on a locally compact abelian group. J. Multivar. Anal. **51**, 1–16 (1994)
8. A. Boudou, Y. Romain: On spectral and random measures associated to discrete and continuous-time processes. Stat. Probab. Lett. **59**, 145–157 (2002)
9. A. Boudou, Y. Romain: On the integral with respect to the tensor product of two random measures. J. Multivar. Anal. **101**(2), 385–394 (2010)
10. S.A. Chobanjan, A. Weron: Banach space-valued stationary processes and their linear prediction. Dissertationes Math. **125**, 1–45 (1975)
11. G. Kallianpur, V. Mandrekar: Spectral Theory of Stationary H-valued Processes. J. Multivar Anal. **1**(1), 1–16 (1971)
12. P. Masani: Orthogonally scattered measures. Adv. Math. **2**, 61–117 (1968)
13. R. Payen: Fonctions aléatoires du second ordre á valeurs dans un espace de Hilbert. Ann. Inst. Henri Poincaré Sect. B **3**(4), 323–396 (1967)
14. I. Singer: Bases in Banach Spaces I. (Springer, Berlin 1970)
15. A.M. Yaglom: An Introduction to the Theory of Stationary Random Function. Dover edition, New York (1987)

On the Strong Consistency of the Kernel Estimator of Extreme Conditional Quantiles

Stéphane Girard and Sana Louhichi

Abstract Nonparametric regression quantiles can be obtained by inverting a kernel estimator of the conditional distribution. The asymptotic properties of this estimator are well known in the case of ordinary quantiles of fixed order. The goal of this paper is to establish the strong consistency of the estimator in case of extreme conditional quantiles. In such a case, the probability of exceeding the quantile tends to zero as the sample size increases, and the extreme conditional quantile is thus located in the distribution tails.

1 Introduction

Quantile regression plays a central role in various statistical studies. In particular, nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution function are extensively investigated in the sample case [4, 27, 29, 30]. Extensions to random fields [1], time series [14], functional data [11], and truncated data [25] are also available. However, all these papers are restricted to conditional quantiles having a fixed order $\alpha \in (0,1)$. In the following, α denotes the conditional probability to be larger than the conditional quantile. Consequently, the above-mentioned asymptotic theories do not apply in the distribution tails, i.e. when $\alpha = \alpha_n \to 0$ or $\alpha_n \to 1$ as the sample size n goes to infinity. Motivating applications include, for instance, environmental studies [16, 28], finance [31], assurance [3], and image analysis [26].

The asymptotics of extreme conditional quantile estimators have been established in a number of regression models. Chernozhukov [6] and Jurecková [22] considered the extreme quantiles in the linear regression model and derived their asymptotic distributions under various error distributions. Other parametric models are considered in [10, 28]. A semi-parametric approach to modeling trends

S. Girard
Mistis, Inria Grenoble Rhône-Alpes and Laboratoire Jean Kuntzmann, Grenoble, France
e-mail: Stephane.Girard@inria.fr

S. Louhichi (✉)
Laboratoire Jean Kuntzmann, Univ. Grenoble-Alpes, Saint-Martin-d'Hères, France
e-mail: Sana.Louhichi@imag.fr

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_4

in extremes has been introduced in [9] basing on local polynomial fitting of the Generalized extreme value distribution. Hall and Tajvidi [21] suggested a nonparametric estimation of the temporal trend when fitting parametric models to extreme-values. Another semi-parametric method has been developed in [2] using a conditional Pareto-type distribution for the response. Fully nonparametric estimators of extreme conditional quantiles have been discussed in [2, 5] including local polynomial maximum likelihood estimation, and spline fitting via maximum penalized likelihood. Recently, [15, 18] proposed, respectively, a moving-window based estimator for the tail index and extreme quantiles of heavy-tailed conditional distributions, and they established their asymptotic properties.

In the kernel-smoothing case, the asymptotic theory for quantile regression in the tails is still in full development. Girard and Jacob [19] and Girard and Menneteau [20] have analyzed the case $\alpha_n = 1/n$ in the particular situation where the response Y given $X = x$ is uniformly distributed. The asymptotic distribution of the kernel estimator of extreme conditional quantile is established by [7, 17] for heavy-tailed conditional distributions. This result is extended to all types of tails in [8].

Here, we focus on the strong consistency of the kernel estimator for extreme conditional quantiles. Our main result is established in Sect. 2. Some illustrative examples are provided in Sect. 3. The proofs of the main results are given in Sect. 4 and the proofs of the auxiliary results are given in the Appendix.

2 Main Results

Let $(X_i, Y_i)_{1\le i\le n}$ be independent copies of a random pair $(X, Y) \in \mathbb{R}^d \times \mathbb{R}^+$ with density $f_{(X,Y)}$. Let g be the density of X that we suppose strictly positive. The conditional survival function of Y given $X = x$ is denoted by:

$$\bar{F}(y|x) = \mathbb{P}(Y > y|X = x) = \frac{1}{g(x)} \int_y^{+\infty} f_{(X,Y)}(x,z)dz.$$

The kernel estimator of $\bar{F}(y|x)$ is, for x such that $\sum_{i=1}^n K_h(x - X_i) \neq 0$:

$$\bar{F}_n(y|x) = \frac{\sum_{i=1}^n K_h(x - X_i)\mathrm{I}_{Y_i>y}}{\sum_{i=1}^n K_h(x - X_i)},$$

where $h = h_n \to 0$ as $n \to \infty$ and $K_h(u) = \frac{1}{h^d} K(\frac{u}{h})$, the kernel K is a measurable function which satisfies the conditions:

($\mathscr{K}_1$) K is a continuous probability density.
($\mathscr{K}_2$) K is with compact support: $\exists\, R > 0$ such that $K(u) = 0$ for any $\|u\| \le R$.
Define $\kappa := \|K\|_\infty = \sup_{x\in\mathbb{R}^d} K(x) < \infty$.

Recall that, for a class of function $\mathscr{G}$, $\mathscr{N}(\epsilon, \mathscr{G}, d_Q)$ denotes the minimal number of balls $\{g, d_Q(g, g') < \epsilon\}$ of d_Q-radius ϵ needed to cover $\mathscr{G}$, and d_Q is the $L_2(Q)$-metric. Let $\mathscr{K}$ be the set of functions $\mathscr{K} = \{K((x-\cdot)/h),\ h > 0, x \in \mathbb{R}^d\}$ and $\mathscr{N}(\epsilon, \mathscr{K}) = \sup_Q \mathscr{N}(\epsilon \|K\|_\infty, \mathscr{K}, d_Q)$, where the supremum is taken over all the probability measure Q on $\mathbb{R}^d \times \mathbb{R}$. Suppose that,

($\mathscr{K}_3$) for some $C, \nu > 1$, $\mathscr{N}(\epsilon, \mathscr{K}) \leq C\epsilon^{-\nu}$ for any $\epsilon \in]0, 1[$.

A number of sufficient conditions for which ($\mathscr{K}_3$) holds are discussed in [13] and the references therein. Finally suppose that

($\mathscr{A}_1$) for all $\alpha \in (0, 1)$ there exists an unique $q(\alpha|x) \in \mathbb{R}$ such that $\bar{F}(q(\alpha|x)|x) = \alpha$.

The conditional quantile $q(\alpha|x)$ is then the inverse of the function $\bar{F}(\cdot|x)$ at the point α. Let (α_n) be a fixed sequence of levels with values in $[0, 1]$. For any $x \in \mathbb{R}^d$, define $q(\alpha_n|x)$ as the unique solution of the equation:

$$\alpha_n = \bar{F}(q(\alpha_n|x)|x), \tag{1}$$

whose existence is guaranteed by Assumption ($\mathscr{A}_1$). The kernel estimator of the conditional quantiles $q(\alpha|x)$ is:

$$\hat{q}_n(\alpha|x) = \inf\{y \in \mathbb{R}, \bar{F}_n(y|x) \leq \alpha\}. \tag{2}$$

Finally, denote by $\hat{g}_n(x)$ the kernel density estimator of the probability density g, i.e.:

$$\hat{g}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i).$$

Our main result is the following.

Theorem 1 *Let $(X_i, Y_i)_{1\leq i\leq n}$ be independent copies of a random pair $(X, Y) \in \mathbb{R}^d \times \mathbb{R}^+$ with density $f_{(X,Y)}$. Let g be the density of X that we suppose bounded and strictly positive. Suppose that Assumption ($\mathscr{A}_1$) is satisfied. Let (α_n) be a sequence of levels in $[0, 1]$ for which*

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\hat{q}_n(\alpha_n|x)}{q(\alpha_n|x)} \leq Cst, \tag{3}$$

almost surely. Suppose that the kernel K satisfies Conditions ($\mathscr{K}_1$), ($\mathscr{K}_2$), and ($\mathscr{K}_3$). Define

$$A(y, z, x, h_n) = \sup_{u: d(u,x)\leq Rh_n} \left| \frac{\bar{F}(y|u)}{\bar{F}(z|x)} - 1 \right|.$$

Suppose that for some fixed positive ϵ_0 *and for* $z \in \{q(\alpha_n|x), (1+\epsilon)q(\alpha_n|x)\}$

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}^d, |\epsilon|\le\epsilon_0} A((1+\epsilon)q(\alpha_n|x), z, x, h_n) \le C < \infty. \tag{4}$$

If, moreover,

$$\lim_{n\to\infty} nh_n^d\alpha_n = \infty \text{ and } \lim_{n\to\infty} \frac{\ln(\alpha_n h_n^d \wedge \alpha_n^2)}{nh_n^d\alpha_n} = 0, \tag{5}$$

then there exists a positive constant C, not dependent on x, such that one has for n sufficiently large

$$\left|1 - \frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)}\right| \le C \sup_{|\epsilon|\le\epsilon_0} A((1+\epsilon)q(\alpha_n|x), (1+\epsilon)q(\alpha_n|x), x, h_n)$$
$$+ \frac{C}{\hat{g}_n(x)} \sqrt{\frac{\ln(\alpha_n^{-1}h_n^{-d}) \vee \ln\ln n}{n\alpha_n h_n^d}}, \text{ almost surely.}$$

The first term of the last bound can be interpreted as a bias term due to the kernel smoothing. The second term can be seen as a variance term, $n\alpha_n h_n^d$ being the effective number of points used in the estimation. The following proposition gives conditions under which (3) is satisfied.

Proposition 1 *Suppose that g is Lipschitzian and bounded above by* $g_{\max}$. *Let* $v_d = \int_{\|v\|\le 1} dv$ *be the volume of the unit sphere. If* $A(q(\alpha_n|x), q(\alpha_n|x), x, 0, h) \to 0$ *as* $n \to \infty$ *and there exist* $\varepsilon > 0$ *such that*

$$\sum_{n=1}^{\infty} nh^d\alpha_n \exp\{-v_d g_{\max} nh_d \alpha_n (1+\varepsilon)\} = \infty, \tag{6}$$

then,

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}^d} \frac{\hat{q}_n(\alpha_n|x)}{q(\alpha_n|x)} \le 1, \text{ almost surely.}$$

Some examples of distributions satisfying condition (4) are provided in the next section.

3 Examples

Let us first focus on a conditional Pareto distribution defined as:

$$\bar{F}(y|x) = y^{-\theta(x)}, \text{ for all } y > 0. \tag{7}$$

Here, $\theta(x) > 0$ can be read as the inverse of the conditional extreme-value index. The above distribution belongs to the so-called Fréchet maximum domain of attraction which encompasses all distributions with heavy tails. As a consequence of Theorem 1, we have:

Corollary 1 *Let us consider a conditional Pareto distribution (7) such that* $0 < \theta_{\min} \leq \theta(x) \leq \theta_{\max}$ *for all* $x \in \mathbb{R}^d$*. Assume that* θ *is Lipschitzian. If the sequences* (α_n) *and* (h_n) *are such that* $h_n \log \alpha_n \to 0$ *as* $n \to \infty$ *and (5), (6) hold, then* $\hat{q}_n(\alpha_n|x)/q(\alpha_n|x) \to 1$ *almost surely as* $n \to \infty$.

Let us now consider a conditional exponential distribution defined as:

$$\bar{F}(y|x) = \exp(-\theta(x)y), \text{ for all } y > 0, \tag{8}$$

where $\theta(x) > 0$ is the inverse of the conditional expectation of Y given $X = x$. This distribution belongs to the Gumbel maximum domain of attraction which collects all distributions with a null conditional extreme-value index. These distributions are often referred to as light-tailed distributions. In such a case, Theorem 1 yields a stronger convergence result than in the heavy-tail framework:

Corollary 2 *Let us consider a conditional exponential distribution (8) with* $0 < \theta_{\min} \leq \theta(x) \leq \theta_{\max}$ *for all* $x \in \mathbb{R}^d$*. Assume that* θ *is Lipschitzian. If the sequences* (α_n) *and* (h_n) *are such that* $h_n \log \alpha_n \to 0$ *as* $n \to \infty$ *and (5), (6) hold, then* $(\hat{q}_n(\alpha_n|x) - q(\alpha_n|x)) \to 0$ *almost surely as* $n \to \infty$.

4 Proofs

4.1 Proof of Theorem 1

Clearly, by (1):

$$\frac{|\bar{F}(q(\alpha_n|x)|x) - \bar{F}(\hat{q}_n(\alpha_n|x)|x)|}{\bar{F}(q(\alpha_n|x)|x)} \leq \frac{|\alpha_n - \bar{F}_n(\hat{q}_n(\alpha_n|x)|x)|}{\bar{F}(q(\alpha_n|x)|x)} + \frac{|\bar{F}_n(\hat{q}_n(\alpha_n|x)|x) - \bar{F}(\hat{q}_n(\alpha_n|x)|x)|}{\bar{F}(q(\alpha_n|x)|x)}.$$

First, from (2), $|\alpha_n - \bar{F}_n(\hat{q}_n(\alpha_n|x)|x)|$ is bounded above by the maximal jump of $\bar{F}_n(y|x)$ at some observation point (X_j, Y_j):

$$|\alpha_n - \bar{F}_n(\hat{q}_n(\alpha_n|x)|x)| \leq \frac{\max_{j=1,\dots,n} K_h(x - X_j)}{\sum_{i=1}^n K_h(x - X_i)}.$$

It follows from $(\mathscr{K}_2)$ that

$$\frac{|\alpha_n - \bar{F}_n(\hat{q}_n(\alpha_n|x)|x)|}{\bar{F}(q(\alpha_n|x)|x)} \leq \frac{\kappa}{nh^d \alpha_n \hat{g}_n(x)}.$$

Let us then focus on the second term:

$$\frac{|\bar{F}_n(\hat{q}_n(\alpha_n|x)|x) - \bar{F}(\hat{q}_n(\alpha_n|x)|x)|}{\bar{F}(q(\alpha_n|x)|x)} = \frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)} \left| \frac{\bar{F}_n(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(\hat{q}_n(\alpha_n|x)|x)} - 1 \right|. \quad (9)$$

We write $\hat{q}_n(\alpha_n|x) = (1 + \epsilon)\, q(\alpha_n|x)$, with $\epsilon = \frac{\hat{q}_n(\alpha_n|x)}{q(\alpha_n|x)} - 1$. Condition (3) allows to deduce that there exists, for n sufficiently large, ϵ_0 not dependent on x such that $|\epsilon| \leq \epsilon_0$. Consequently and taking into account (9) there exists a positive constant ϵ_0 not dependent on x and n such that (for the sake of simplicity, we write $q = q(\alpha_n|x)$):

$$\left| \frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)} - 1 \right| \leq \frac{\kappa}{nh^d \alpha_n \hat{g}_n(x)} + \sup_{|\epsilon| \leq \epsilon_0} \left(\frac{\bar{F}(q(1+\epsilon))|x)}{\bar{F}(q|x)} \left| \frac{\bar{F}_n(q(1+\epsilon)|x)}{\bar{F}(q(1+\epsilon))|x)} - 1 \right| \right). \quad (10)$$

Our purpose now is to control the term $\left| \frac{\bar{F}_n(q(1+\epsilon)|x)}{\bar{F}(q(1+\epsilon)|x)} - 1 \right|$ of (10). For this, write:

$$\bar{F}_n(y|x) = \frac{\sum_{i=1}^n K_h(x - X_i) \mathrm{I}_{Y_i > y}}{\sum_{i=1}^n K_h(x - X_i)} =: \frac{\hat{\psi}_n(y, x)}{\hat{g}_n(x)},$$

with

$$\hat{\psi}_n(y, x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \mathrm{I}_{Y_i > y}, \quad \hat{g}_n(x) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i).$$

We need the following lemma.

Lemma 1 *Suppose that Condition ($\mathscr{K}_2$) holds. Then, for each n, one has for any $y \in \mathbb{R}$ and $x \in \mathbb{R}^d$ for which $\bar{F}(y|x)\hat{g}_n(x) \neq 0$:*

$$\left|\frac{\bar{F}_n(y|x)}{\bar{F}(y|x)} - 1\right| \leq A(y, y, x, h_n) + (1 + A(y, y, x, h_n))\frac{|\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)|}{\hat{g}_n(x)}$$
$$+\frac{\left|\hat{\psi}_n(y, x) - \mathbb{E}\left(\hat{\psi}_n(y, x)\right)\right|}{\bar{F}(y|x)\hat{g}_n(x)}.$$

The proof is given in the Appendix. According to Lemma 1, we have to control the two quantities $\mathbb{E}\,|\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)|$ and $\frac{|\hat{\psi}_n(y,x) - \mathbb{E}(\hat{\psi}_n(y,x))|}{\bar{F}(y|x)}$. This is the purpose of Propositions 2 and 3 below.

Proposition 2 (Einmahl–Mason [13]) *Suppose that g is a bounded density on $\mathbb{R}^d$, and that the assumptions $(K.i), \ldots, (K.iv)$ of [13] are all satisfied. Then, for any $c > 0$:*

$$\limsup_{n\to\infty} \sup_{\{c \ln n/n \leq h_n^d \leq 1\}} \sqrt{\frac{nh_n^d}{\ln(h_n^{-d}) \vee \ln\ln n}} \sup_{x\in\mathbb{R}^d} |\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)| =: K(c) < \infty,$$

almost surely.

Our task now is to control $\frac{|\hat{\psi}_n(y,x) - \mathbb{E}(\hat{\psi}_n(y,x))|}{\bar{F}(y|x)}$. Let ϵ be a fixed real in $[-\epsilon_0, \epsilon_0]$ for some arbitrary positive ϵ_0. The following proposition evaluates the almost sure asymptotic behavior of:

$$\frac{|\hat{\psi}_n(q(1+\epsilon), x) - \mathbb{E}(\hat{\psi}_n(q(1+\epsilon), x))|}{\bar{F}(q|x)}.$$

Proposition 3 *Let (α_n) be a sequence in $[0, 1]$ and for $x \in \mathbb{R}^d$, $q = q(\alpha_n|x)$ be the conditional quantile as defined by (1). Define the set of functions $\mathscr{F}$ by:*

$$\mathscr{F} = \left\{(u, v) \longmapsto K\left(\frac{x-u}{h}\right) I_{v>q(1+\epsilon)},\ n \in \mathbb{N},\ h > 0,\ x \in \mathbb{R}^d,\ |\epsilon| \leq \epsilon_0\right\} \tag{11}$$

and suppose that $\mathscr{N}(\epsilon, \mathscr{F}) \leq C\epsilon^{-\nu}$, for some $C, \nu > 1$ and all $\epsilon \in]0, 1[$. Suppose also that Condition (4) is satisfied. If $nh_n^d\alpha_n \to \infty$ and $\frac{\ln(\alpha_n h_n^d \wedge \alpha_n^2)}{nh_n^d\alpha_n} \to 0$ as $n \to \infty$, then there exists a positive constant C_1 such that:

$$\limsup_{n\to\infty} \sqrt{\frac{n\alpha_n h_n^d}{\ln(\alpha_n^{-1}h_n^{-d}) \vee \ln\ln n}} \sup_{x\in\mathbb{R}^d, |\epsilon|\leq\epsilon_0} \frac{|\hat{\psi}_n(q(1+\epsilon), x) - \mathbb{E}(\hat{\psi}_n(q(1+\epsilon), x))|}{\bar{F}(q|x)} \leq C_1,$$

almost surely.

Proof of Proposition 3 We have,

$$\frac{1}{\bar{F}(q|x)}\left(\hat{\psi}_n(q(1+\epsilon),x) - \mathbb{E}\left(\hat{\psi}_n(q(1+\epsilon),x)\right)\right)$$
$$= \frac{1}{n\bar{F}(q|x)}\sum_{i=1}^{n}[K_h(x-X_i)\mathrm{I}_{Y_i>q(1+\epsilon)} - \mathbb{E}(K_h(x-X_i)\mathrm{I}_{Y_i>q(1+\epsilon)})]$$
$$= \frac{1}{nh_n^d}\sum_{i=1}^{n}(v_{h,x,\epsilon}(X_i,Y_i) - \mathbb{E}(v_{h,x,\epsilon}(X_i,Y_i))) = \frac{1}{h_n^d\sqrt{n}}\beta_n(v_{h,x,\epsilon}),$$

where,

$$v_{h,x,\epsilon}(u,v) = K\left(\frac{x-u}{h}\right)\frac{\mathrm{I}_{v>q(1+\epsilon)}}{\bar{F}(q|x)}, \quad \beta_n(g) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(X_i,Y_i) - \mathbb{E}(g(X_i,Y_i))).$$

Define the class of functions:

$$\mathscr{G} := \mathscr{G}_{n,h} = \{v_{h,x,\epsilon},\ x \in \mathbb{R}^d, \epsilon \in [-\epsilon_0,\epsilon_0]\} \tag{12}$$

and let $\|\beta_n\|_{\mathscr{G}} = \sup_{g\in\mathscr{G}}|\beta_n(g)|$ and

$$\Theta_n = \sup_{x\in\mathbb{R}^d,\epsilon\in[-\epsilon_0,\epsilon_0]}\frac{\left|\hat{\psi}_n(q(1+\epsilon),x) - \mathbb{E}\left(\hat{\psi}_n(q(1+\epsilon),x)\right)\right|}{\bar{F}(q|x)}.$$

Consequently, for any $\gamma > 0$:

$$\mathbb{P}\left(\Theta_n > \gamma\right) \le \mathbb{P}\left(\sqrt{n}\|\beta_n\|_{\mathscr{G}} > \gamma nh^d\right) \le \mathbb{P}\left(\max_{1\le m\le n}\sqrt{m}\|\beta_m\|_{\mathscr{G}} > \gamma nh^d\right). \tag{13}$$

We then have to evaluate $\max_{1\le m\le n}\sqrt{m}\|\beta_m\|_{\mathscr{G}}$. By Talagrand Inequality (see A.1. in [12]), we have for any $t > 0$ and suitable finite constants $A_1, A_2 > 0$:

$$\mathbb{P}\left(\max_{1\le m\le n}\sqrt{m}\|\beta_m\|_{\mathscr{G}} > A_1\left(\mathbb{E}\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\|_{\mathscr{G}} + t\right)\right)$$
$$\le 2\exp(-A_2t^2/n\sigma^2) + 2\exp(-A_2t/M),$$

where $(\epsilon_i)_i$ is a sequence of independent Rademacher random variables independent of the random vectors $(X_i,Y_i)_{1\le i\le n}$ and

$$\sup_{g\in\mathscr{G}}\|g\|_\infty \le M, \quad \sup_{g\in\mathscr{G}}\mathrm{Var}(g(X,Y)) \le \sigma^2.$$

Here $\|g\|_\infty \le \frac{\|K\|_\infty}{\alpha_n} =: M$, and

$$\begin{aligned}
\mathrm{Var}(v^2_{h,x,\epsilon}(X,Y)) &\le \mathbb{E}(v^2_{h,x,\epsilon}(X,Y)) = \frac{1}{\bar{F}^2(q|x)}\mathbb{E}\left(K^2\left(\frac{x-X}{h}\right)\mathrm{i}_{Y>q(1+\epsilon)}\right) \\
&= \frac{1}{\bar{F}(q|x)}\int K^2\left(\frac{x-u}{h}\right)\frac{\bar{F}(q(1+\epsilon)|u)}{\bar{F}(q|x)}g(u)du \\
&\le \frac{h_n^d}{\alpha_n}\sup_{x\in\mathbb{R}^d,\epsilon\in[-\epsilon_0,\epsilon_0]}(1+A(q(1+\epsilon),q,x,h))\|K\|_2^2\|g\|_\infty \\
&= \frac{h_n^d}{\alpha_n}L =: \sigma^2, \qquad (14)
\end{aligned}$$

for some positive constant L, since for n sufficiently large

$$\sup_{x\in\mathbb{R}^d,\,\epsilon\in[-\epsilon_0,\epsilon_0]} A(q(1+\epsilon),q,x,h) \le cst.$$

We obtain combining this with the above Talagrand's Inequality:

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le m\le n}\sqrt{m}\|\beta_m\|_\infty > A_1\left(\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} + t\right)\right) \\
&\le 2\exp\left(-A_2 t^2\frac{\alpha_n}{nh_n^d L}\right) + 2\exp\left(-A_2 t\frac{\alpha_n}{\|K\|_\infty}\right). \qquad (15)
\end{aligned}$$

The last bound together with (13) give

$$\begin{aligned}
&\mathbb{P}\left(\Theta_n > A_1 n^{-1}h_n^{-d}\left(\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} + t\right)\right) \\
&\le 2\exp\left(-A_2 t^2\frac{\alpha_n}{nh_n^d L}\right) + 2\exp\left(-A_2 t\frac{\alpha_n}{\|K\|_\infty}\right). \qquad (16)
\end{aligned}$$

We now have to evaluate $\mathbb{E}\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\|_{\mathscr{G}}$. We will argue as for the proof of Proposition A.1. in [12]. We have by (6.9) of Proposition 6.8 in [24],

$$\mathbb{E}\left\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} \le 6t_n + 6\mathbb{E}(\max_{i\le n}\|\epsilon_i g(X_i,Y_i)\|_{\mathscr{G}}) \le 6t_n + 6\frac{\|K\|_\infty}{\alpha_n}, \quad (17)$$

where we have defined

$$t_n = \inf\left\{t>0, \mathbb{P}\left(\left\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} > t\right) \le \frac{1}{24}\right\}.$$

Our purpose is then to control t_n. Define the event

$$F_n = \left\{ n^{-1} \sup_{g \in \mathscr{G}} \sum_{j=1}^{n} g^2(X_j, Y_j) \leq 64\sigma^2 \right\},$$

where σ^2 is as in (14). Let g_0 be a fixed element of $\mathscr{G}$. We have,

$$\mathbb{E} \left| \sum_{i=1}^{n} \epsilon_i g_0(X_i, Y_i) \mathrm{I}_{F_n} \right| \leq 8\sigma\sqrt{n}.$$

By (A8) of Einmahl and Mason [12], we now have to control $\mathscr{N}(\epsilon, \mathscr{G}, d_{n,2})$. Recall that $\mathscr{N}(\epsilon, \mathscr{G}, d_Q)$ is the minimal number of balls $\{g, d_Q(g, g') < \epsilon\}$ of d_Q-radius ϵ needed to cover $\mathscr{G}$, d_Q is the $L_2(Q)$-metric and

$$d_{n,2}(f, g) = d_{Q_n}(f, g) = \int (f(x) - g(x))^2 dQ_n(x),$$

with $Q_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{(X_i, Y_i)}$. In other words

$$d_{n,2}(f, g) = \frac{1}{n} \sum_{i=1}^{n} (f(X_i, Y_i) - g(X_i, Y_i))^2.$$

We first note that, on the event F_n,

$$\mathscr{N}(\epsilon, \mathscr{G}, d_{n,2}) = 1, \quad \text{whenever } \epsilon > 16\sigma.$$

We will suppose then $\epsilon \leq 16\sigma$. We have

$$\mathscr{N}(\epsilon, \mathscr{G}, d_{n,2}) = \mathscr{N}(\epsilon, \mathscr{G}, d_{Q_n}) \leq \mathscr{N}(\epsilon\alpha_n, \mathscr{F}, d_{Q_n}),$$

where $\mathscr{F}$ is as defined by (11). Recall that $\mathscr{N}(\epsilon, \mathscr{F}) = \sup_Q \mathscr{N}(\epsilon\|K\|_\infty, \mathscr{F}, d_Q)$ where the supremum is taken over all the probability measure Q on $\mathbb{R}^d \times \mathbb{R}$. We have supposed that:

$$\mathscr{N}(\epsilon, \mathscr{F}) \leq C\epsilon^{-\nu}.$$

for some $C, \nu > 1$ and all $\epsilon \in]0, 1[$. Consequently,

$$\mathscr{N}(\epsilon, \mathscr{G}, d_{n,2}) \leq C \left(\frac{\alpha_n \epsilon}{\|K\|_\infty} \right)^{-\nu}, \tag{18}$$

as soon as $\alpha_n\epsilon < \|K\|_\infty$. Hence, we have almost surely on the event F_n,

$$\begin{aligned}\int_0^\infty \sqrt{\ln(\mathcal{N}(\epsilon,\mathcal{G},d_{n,2}))}d\epsilon &= \int_0^{16\sigma} \sqrt{\ln(\mathcal{N}(\epsilon,\mathcal{G},d_{n,2}))}d\epsilon \\ &\leq \sum_{i=0}^\infty \int_{2^{-i-1}16\sigma}^{2^{-i}16\sigma} \sqrt{\ln(C)+\nu\ln\left(\frac{\|K\|_\infty}{\alpha_n\epsilon}\right)}d\epsilon \\ &\leq \sum_{i=0}^\infty 2^{-i-1}16\sigma\sqrt{\ln(C)+\nu\ln\left(\frac{2^{i+1}\|K\|_\infty}{\alpha_n 16\sigma}\right)} \\ &\leq C_2\sum_{i=0}^\infty \frac{\sqrt{i+1}}{2^{i+1}}\sqrt{\sigma^2\max\left(\ln(C_1),\ln\left(\frac{1}{\alpha_n^2\sigma^2}\right)\right)},\end{aligned}$$

for some positive constants C_1, C_2 that depend only on C, ν and $\|K\|_\infty$. We conclude, using (A8) of [12]:

$$\mathbb{E}\left(\left\|\sum_{i=1}^n \epsilon_i g(X_i,Y_i)\right\|_{\mathcal{G}} \mathrm{i}_{F_n}\right) \leq C_2'\sqrt{n\sigma^2\max\left(\ln(C_1),\ln\left(\frac{1}{\alpha_n^2\sigma^2}\right)\right)}. \tag{19}$$

We now use Inequality A2 in [12] (which is due to Giné and Zinn), with $t = 64\sqrt{n}\sigma^2$. We obtain, for $m \geq 1$, since for any $g \in \mathcal{G}$, $\|g\|_\infty \leq \frac{\|K\|_\infty}{\alpha_n}$,

$$\begin{aligned}\mathbb{P}(F_n^c) &= \mathbb{P}\left(n^{-1}\sup_{g\in\mathcal{G}}\sum_{j=1}^n g^2(X_i,Y_i) > 64\sigma^2\right) \\ &\leq 4\mathbb{P}(\mathcal{N}(\rho n^{-1/4},\mathcal{G},d_{n,2}) \geq m) + 8m\exp(-n\sigma^2\alpha_n^2/\|K\|_\infty^2),\end{aligned}$$

where $n^{-1/4}\rho = n^{-1/4}\min(\sigma n^{1/4}, n^{1/4}) = \min(\sigma,1)$. Hence by (18):

$$\mathcal{N}(\rho n^{-1/4},\mathcal{G},d_{n,2}) \leq C\left(\frac{\alpha_n\min(\sigma,1)}{\|K\|_\infty}\right)^{-\nu},$$

Consequently, we have for $m = [2C(\frac{\alpha_n\min(\sigma,1)}{\|K\|_\infty})^{-\nu}]$:

$$\mathbb{P}(F_n^c) \leq 16C\left(\frac{\alpha_n\min(\sigma,1)}{\|K\|_\infty}\right)^{-\nu}\exp(-n\sigma^2\alpha_n^2/\|K\|_\infty^2).$$

The last bound together with (19) gives:

$$
\begin{aligned}
\mathbb{P}\left(\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} > t\right) &\leq \mathbb{P}(F_n^c) + \frac{1}{t}\mathbb{E}\left(\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} \mathrm{i}_{F_n}\right)\\
&\leq 16C\left(\frac{\|K\|_\infty}{\alpha_n\min(\sigma,1)}\right)^{\nu}\exp(-n\sigma^2\alpha_n^2/\|K\|_\infty^2)\\
&\quad + \frac{C_2'}{t}\sqrt{n\sigma^2\max\left(\ln(C_1),\ln\left(\frac{1}{\alpha_n^2\sigma^2}\right)\right)}. \qquad (20)
\end{aligned}
$$

Let us control the second term in (20). Recall that by (14), $\sigma^2 = L\frac{h_n^d}{\alpha_n}$ and thus $\alpha_n^2\sigma^2 = Lh_n^d\alpha_n$ for some positive constant L. This fact together with $h_n^d\alpha_n \to 0$ as $n\to\infty$ allows us to deduce that for n sufficiently large:

$$
\frac{C_2'}{t}\sqrt{n\sigma^2\max\left(\ln(C_1),\ln\left(\frac{1}{\alpha_n^2\sigma^2}\right)\right)} \leq \frac{Cst}{t}\sqrt{\frac{nh_n^d}{\alpha_n}\ln\left(\frac{1}{\alpha_n h_n^d}\right)}. \qquad (21)
$$

Our task now is to control the first term in (20). We have:

$$
\begin{aligned}
&16C\left(\frac{\|K\|_\infty}{\alpha_n\min(\sigma,1)}\right)^{\nu}\exp(-n\sigma^2\alpha_n^2/\|K\|_\infty^2)\\
&\leq L_0^{\nu/2}\exp\left(-nh_n^d\alpha_n\left(L_1+\frac{\nu}{2nh_n^d\alpha_n}\ln(\alpha_n h_n^d\wedge\alpha_n^2)\right)\right),
\end{aligned}
$$

which tends to 0 as $n\to\infty$ as soon as $nh_n^d\alpha_n\to 0$ and $\frac{\ln(\alpha_n h_n^d\wedge\alpha_n^2)}{nh_n^d\alpha_n}\to 0$. Hence for $t\geq 48Cst\sqrt{\frac{nh_n^d}{\alpha_n}\ln(\frac{1}{\alpha_n h_n^d})} =: t_n$, Inequalities (20) and (21) give for n sufficiently large and for $t\geq t_n$

$$
\mathbb{P}\left(\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} > t\right) \leq \frac{1}{24}.
$$

We conclude using this fact together with Inequality (17):

$$
\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathscr{G}} \leq c\left(\frac{1}{\alpha_n}+\sqrt{\frac{nh_n^d}{\alpha_n}\ln\left(\frac{1}{\alpha_n h_n^d}\right)}\right) = \mathscr{O}\left(\sqrt{\frac{nh_n^d}{\alpha_n}\ln\left(\frac{1}{\alpha_n h_n^d}\right)}\right). \qquad (22)
$$

Recalling that

$$\Theta_n = \sup_{x\in\mathbb{R}^d,\epsilon\in[-\epsilon_0,\epsilon_0]} \frac{\left|\hat{\psi}_n(q(1+\epsilon),x) - \mathbb{E}\left(\hat{\psi}_n(q(1+\epsilon),x)\right)\right|}{\bar{F}(q|x)},$$

and collecting (22), (16), yield

$$\mathbb{P}\left(\Theta_n > A_1 n^{-1} h_n^{-d}\left(\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathcal{G}} + t\right)\right) \leq 2\left[\exp\left(-A_2\tilde{L}^2\frac{\ln(\ln(n))}{L}\right) + \exp\left(-\frac{A_2\tilde{L}}{\|K\|_\infty}\sqrt{nh_n^d\alpha_n\ln(\ln n)}\right)\right], \quad (23)$$

for any $t \geq \left(Cst\sqrt{\frac{nh_n^d}{\alpha_n}\ln(\frac{1}{\alpha_n h_n^d})}\right) \vee \tilde{L}\sqrt{\frac{nh_n^d}{\alpha_n}\ln(\ln n)}$ and some $\tilde{L} > 0$. Now,

$$\frac{\ln n}{n\alpha_n h_n^d} + \frac{\ln(\alpha_n h_n^d \wedge \alpha_n^2)}{n\alpha_n h_n^d} \leq \frac{\ln(n\alpha_n h_n^d)}{n\alpha_n h_n^d},$$

which tends to 0 as n tends to infinity, since $\lim_{n\to\infty} n\alpha_n h_n^d = \infty$. Hence,

$$\lim_{n\to\infty}\frac{\ln n}{n\alpha_n h_n^d} = 0,$$

which proves that for n sufficiently large $nh_n^d\alpha_n \geq \ln n$. We conclude then from (23) that:

$$\mathbb{P}\left(\Theta_n > A_1 n^{-1} h_n^{-d}\left(\mathbb{E}\left\|\sum_{i=1}^{n}\epsilon_i g(X_i,Y_i)\right\|_{\mathcal{G}} + t\right)\right) \leq 4(\ln n)^{-\rho},$$

since we have $nh_n^d\alpha_n \geq \ln n$. We choose $\tilde{L}$ in such a way that $\rho > 1$. Proposition 3 is proved thanks to Borel–Cantelli lemma. ■

We continue the proof of Theorem 1. Inequality (10), together with Lemma 1 and Condition (4), gives for some universal positive constant C:

$$\left|\frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)} - 1\right| \leq \frac{\kappa}{nh^d\alpha_n\hat{g}_n(x)} + C\sup_{|\epsilon|\leq\epsilon_0} A(q(1+\epsilon),q(1+\epsilon),x,h_n)$$
$$+C\sup_{|\epsilon|\leq\epsilon_0}(1 + A(q(1+\epsilon),q(1+\epsilon),x,h_n))\frac{|\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)|}{\hat{g}_n(x)}$$
$$+C\sup_{|\epsilon|\leq\epsilon_0}\frac{\left|\hat{\psi}_n(q(1+\epsilon),x) - \mathbb{E}\left(\hat{\psi}_n(q(1+\epsilon),x)\right)\right|}{\bar{F}(q|x)\hat{g}_n(x)}. \quad (24)$$

We first use Einmahl and Mason's result (cf. Proposition 2 above). All the requirements of Einmahl and Mason's result are satisfied from that of Theorem 1. This gives that, for $c \ln n/n \le h_n^d \le 1$:

$$\sqrt{\frac{nh_n^d}{\ln(h_n^{-d}) \vee \ln\ln n}} \sup_{x\in\mathbb{R}^d} |\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)| < C, \tag{25}$$

almost surely. Our task now is to apply Proposition 3. We first claim that

Lemma 2 *Under Condition* ($\mathscr{K}_3$), *the class of function* $\mathscr{F}$ *defined by* (11) *satisfies* $\mathscr{N}(\epsilon, \mathscr{F}) \le C\epsilon^{-\nu}$, *for* $C > 0, \epsilon > 0, \nu > 1$.

Proof of Lemma 2 Define the set of function $\mathscr{F} = \mathscr{K}\mathscr{I}$, where the set of functions $\mathscr{K}$ is $\mathscr{K} = \left\{u \longmapsto K\left(\frac{x-u}{h}\right), x \in \mathbb{R}^d,\ h > 0\right\}$, and $\mathscr{I} = \{v \longmapsto \mathrm{I}_{v>q(1+\epsilon)},\ x \in \mathbb{R}^d, n \in \mathbb{N}, |\epsilon| \le \epsilon_0\}$. The proof of Lemma 2 follows from Lemma A.1 of [12] since $\mathscr{N}(\epsilon, \{v \longmapsto \mathrm{I}_{v>y},\ y \in \mathbb{R}\}) \le C\epsilon^{-\tilde{\nu}}$ with $\tilde{\nu} > 0$ and $C > 0$. ■

Consequently, all the requirements of Proposition 3 are satisfied from that of Theorem 1. The conclusion of Proposition 3 together with (25), (24) and the facts that:

$$\sqrt{\frac{\ln(h_n^{-d}) \vee \ln\ln n}{nh_n^d}} \le \sqrt{\frac{\ln(\alpha_n^{-1}h_n^{-d}) \vee \ln\ln n}{nh_n^d\alpha_n}},$$

$$\frac{1}{nh_n^d\alpha_n} \le \sqrt{\frac{\ln(\alpha_n^{-1}h_n^{-d}) \vee \ln\ln n}{n\alpha_n h_n^d}}$$

complete the proof of Theorem 1. ■

4.2 Proof of Proposition 1

Let us introduce $Z_i^{(n)}(x)$ for $i = 1, \ldots, n$ a triangular array of i.i.d. random variables defined by $Z_i^{(n)}(x) = Y_i \mathrm{i}_{\|x-X_i\| \le h}$. Their common survival distribution function can be expanded as:

$$\begin{aligned}\bar{\Psi}_n(t, x) = \mathbb{P}(Z_1^{(n)}(x) > t) &= \int_{\|x-u\|\le h} \bar{F}(t|u)g(u)du \\ &= h^d \int_{\|v\|\le 1} \bar{F}(t|x-hv)g(x-hv)dv,\end{aligned}$$

or equivalently,

$$\frac{\bar{\Psi}_n(t,x)}{h^d \bar{F}(t|x)g(x)} = \int_{\|v\|\leq 1} dv + \int_{\|v\|\leq 1} \left(\frac{\bar{F}(t|x-hv)}{\bar{F}(t|x)} - 1 \right) \frac{g(x-hv)}{g(x)} dv$$
$$+ \int_{\|v\|\leq 1} \left(\frac{g(x-hv)}{g(x)} - 1 \right) dv.$$

Letting $v_d = \int_{\|v\|\leq 1} dv$ the volume of the unit sphere, and assuming that g is Lipschitzian, it follows,

$$\frac{\bar{\Psi}_n(t,x)}{h^d \bar{F}(t|x)g(x)} = v_d + o(1) + O(A(t,t,x,0,h))$$

and introducing $\beta_n(x) = n\bar{\Psi}_n(q(\alpha_n|x)|x)$, we obtain

$$\beta_n(x) = v_d g(x) n h_d \alpha_n (1 + o(1))$$

under condition $A(q(\alpha_n|x), q(\alpha_n|x), x, 0, h) \to 0$ as $n \to \infty$. We now need the following lemma (also available to triangular arrays).

Lemma 3 (Klass [23]) *Let $Z, Z_1, Z_2, \ldots$ be a sequence of i.i.d. random vectors and define $M_n = \max\{Z_1, \ldots, Z_n\}$. Suppose that (b_n) is nondecreasing, $\mathbb{P}(Z > b_n) \to 0$ and $n\mathbb{P}(Z > b_n) \to \infty$ as $n \to \infty$. If, moreover,*

$$\sum_{n=1}^{\infty} \mathbb{P}(Z > b_n) \exp\{-n\mathbb{P}(Z > b_n)\} = \infty,$$

then $\limsup_{n\to\infty} M_n/b_n < 1$ *a.s.*

From Lemma 3, a sufficient condition for:

$$\limsup_{n\to\infty} \frac{\max_{1\leq i\leq n} Z_i^{(n)}(x)}{q(\alpha_n|x)} < 1 \text{ a.s.} \tag{26}$$

is

$$\sum_{n=1}^{\infty} \frac{\beta_n(x)}{n} \exp\{-\beta_n(x)\} = \infty,$$

which is fulfilled under (6). Finally,

$$\frac{\hat{q}_n(\alpha_n|x)}{q(\alpha_n|x)} \leq \frac{\max_{1\leq i\leq n} Z_i^{(n)}(x)}{q(\alpha_n|x)}$$

and the conclusion follows from (26). ∎

4.3 Proofs of Corollaries

Proof of Corollary 1 For all $\tau \in [0, 1]$ and (u, x) such that $d(u, x) \le Rh_n$, we have

$$
\begin{aligned}
\frac{\bar{F}(q(\alpha_n|x)(1+\varepsilon)|u)}{\bar{F}(q(\alpha_n|x)(1+\varepsilon)^\tau|x)} &= q(\alpha_n|x)^{\theta(x)-\theta(u)}(1+\varepsilon)^{\tau\theta(x)-\theta(u)} \\
&= \exp\left\{\frac{\theta(u)-\theta(x)}{\theta(x)}\log\alpha_n + (\tau\theta(x)-\theta(u))\log(1+\varepsilon)\right\} \\
&= \exp\{O(h_n\log\alpha_n) + O(\tau\theta(x)-\theta(u))\} \\
&= (1+O(h_n\log\alpha_n))\exp\{O(\tau\theta(x)-\theta(u))\}. \qquad (27)
\end{aligned}
$$

Assuming that $h_n \log\alpha_n \to 0$ as $n \to \infty$, it follows that (27) is bounded above for all $\tau \in [0, 1]$ and (u, x) such that $d(u, x) \le Rh_n$ and therefore condition (4) is fulfilled. If, moreover, $\tau = 1$ then $O(\theta(x) - \theta(u)) = O(h_n)$, and thus (27) tends to zero as n goes to infinity. Proposition 1 then entails that assumption (3) holds. Theorem 1 implies that:

$$
\left|1 - \frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)}\right| = \left|1 - \left(\frac{\hat{q}_n(\alpha_n|x)}{q(\alpha_n|x)}\right)^{-\theta(x)}\right| \to 0
$$

almost surely as $n \to \infty$. The conclusion follows. ■

Proof of Corollary 2 For all $\tau \in [0, 1]$ and (u, x) such that $d(u, x) \le Rh_n$, we have

$$
\begin{aligned}
\frac{\bar{F}(q(\alpha_n|x)(1+\varepsilon)|u)}{\bar{F}(q(\alpha_n|x)(1+\varepsilon)^\tau|x)} &= \exp\left\{(1+\varepsilon)\log(\alpha_n)\left(\frac{\theta(u)-\theta(x)}{\theta(x)} + 1 - (1+\varepsilon)^{\tau-1}\right)\right\} \\
&= \exp\{O(h_n\log\alpha_n)\}\exp\left\{(1+\varepsilon)\log(\alpha_n)\left(1-(1+\varepsilon)^{\tau-1}\right)\right\} \\
&= (1+O(h_n\log\alpha_n)o(1) = o(1). \qquad (28)
\end{aligned}
$$

Assuming that $h_n \log\alpha_n \to 0$ as $n \to \infty$, it follows that (28) tends to zero as n goes to infinity. Assumptions (3) and (4) both hold. Theorem 1 implies that:

$$
\left|1 - \frac{\bar{F}(\hat{q}_n(\alpha_n|x)|x)}{\bar{F}(q(\alpha_n|x)|x)}\right| = |1 - \exp((q(\alpha_n|x) - \hat{q}_n(\alpha_n|x))\theta(x))| \to 0
$$

almost surely as $n \to \infty$. The conclusion follows. ■

Appendix: Proof of Auxiliary Results

Proof of Lemma 1 Clearly:

$$\left|\frac{\bar{F}_n(y|x)}{\bar{F}(y|x)}-1\right| \leq \left|\frac{\bar{F}_n(y|x)}{\bar{F}(y|x)}-\frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)}{\bar{F}(y|x)\mathbb{E}\left(\hat{g}_n(x)\right)}\right| + \left|\frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)}{\bar{F}(y|x)\mathbb{E}\left(\hat{g}_n(x)\right)}-1\right|. \tag{29}$$

We have,

$$\begin{aligned}
\mathbb{E}\left(\hat{\psi}_n(y,x)\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(K_h(x-X_i)\mathrm{i}_{Y_i>y}) = \mathbb{E}(K_h(x-X_1)\mathrm{i}_{Y_1>y}) \\
&= \mathbb{E}(K_h(x-X_1)\mathbb{P}(Y_1>y|X_1)) = \int K_h(x-z)\mathbb{P}(Y_1>y|X_1=z)g(z)dz \\
&= \int K_h(x-z)\bar{F}(y|z)g(z)dz,
\end{aligned}$$

and $\mathbb{E}\left(\hat{g}_n(x)\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(K_h(x-X_i)) = \mathbb{E}(K_h(x-X_1))$. Consequently,

$$\begin{aligned}
\frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)}{\bar{F}(y|x)\mathbb{E}\left(\hat{g}_n(x)\right)}-1 &= \frac{1}{\mathbb{E}(K_h(x-X_1))}\left(\int K_h(x-z)\left[\frac{\bar{F}(y|z)}{\bar{F}(y|x)}-1\right]g(z)dz\right) \\
&= \frac{1}{\mathbb{E}(K_h(x-X_1))}\left(\int K_h(u)\left[\frac{\bar{F}(y|x-u)}{\bar{F}(y|x)}-1\right]g(x-u)du\right).
\end{aligned}$$

We conclude, since the kernel K is compactly supported, that for some $R>0$,

$$\left|\frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)}{\bar{F}(y|x)\mathbb{E}\left(\hat{g}_n(x)\right)}-1\right| \leq \sup_{\{x',d(x,x')\leq hR\}}\left|\frac{\bar{F}(y|x')}{\bar{F}(y|x)}-1\right| = A(y,y,x,h). \tag{30}$$

Now,

$$\begin{aligned}
&\left|\frac{\bar{F}_n(y|x)}{\bar{F}(y|x)}-\frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)}{\bar{F}(y|x)\mathbb{E}\left(\hat{g}_n(x)\right)}\right| \\
&\quad\leq \frac{\left|\hat{\psi}_n(y,x)-\mathbb{E}\left(\hat{\psi}_n(y,x)\right)\right|}{\bar{F}(y|x)\hat{g}_n(x)} + \frac{\mathbb{E}\left(\hat{\psi}_n(y,x)\right)|\hat{g}_n(x)-\mathbb{E}\hat{g}_n(x)|}{\bar{F}(y|x)\hat{g}_n(x)\mathbb{E}\hat{g}_n(x)}
\end{aligned}$$

$$\leq \frac{\left|\hat{\psi}_n(y,x) - \mathbb{E}\left(\hat{\psi}_n(y,x)\right)\right|}{\bar{F}(y|x)\hat{g}_n(x)} + (1 + A(y,y,x,h))\frac{\mathbb{E}\,|\hat{g}_n(x) - \mathbb{E}\hat{g}_n(x)|}{\hat{g}_n(x)},$$

by (30). The last bound together with (30) and (29) prove Lemma 1. ■

References

1. Abdi, S., Abdi, A., Dabo-Niang, S., Diop, A.: Consistency of a nonparametric conditional quantile estimator for random fields. Math. Methods Stat. **19**, 1–21 (2010)
2. Beirlant, J., Goegebeur, Y.: Local polynomial maximum likelihood estimation for Pareto-type distributions. J. Multivar. Anal. **89**, 97–118 (2004)
3. Beirlant, J., Goegebeur, Y., Segers, J., Teugels J.: Statistics of Extremes: Theory and Applications. Wiley, New York (2004)
4. Berlinet, A., Gannoun, A., Matzner-Lober, E.: Asymptotic normality of convergent estimates of conditional quantiles. Statistics **35**, 139–169 (2001)
5. Chavez-Demoulin, V., Davison, A.C.: Generalized additive modelling of sample extremes. J. R. Stat. Soc. Ser. C **54**, 207–222 (2005)
6. Chernozhukov, V.: Extremal quantile regression. Ann. Stat. **33**, 806–839 (2005)
7. Daouia, A., Gardes, L., Girard, S., Lekina, A.: Kernel estimators of extreme level curves. Test **20**, 311–333 (2011)
8. Daouia, A., Gardes, L., Girard, S.: On kernel smoothing for extremal quantile regression. Bernoulli **19**, 2557–2589 (2013)
9. Davison, A.C., Ramesh, N.I.: Local likelihood smoothing of sample extremes. J. R. Stat. Soc. Ser. B **62**, 191–208 (2000)
10. Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds. J. R. Stat. Soc. Ser. B **52**, 393–442 (1990)
11. Dabo-Niang, S., Laksaci, A.: Nonparametric quantile regression estimation for functional dependent data. Commun. Stat. Theory Methods **41**, 1254–1268 (2012)
12. Einmahl, U., Mason, D.M.: An empirical process approach to the uniform consistency of kernel type function estimators. J. Theor. Probab. **13**, 1–37 (2000)
13. Einmahl, U., Mason, D.M.: Uniform in bandwidth consistency of kernel type function estimators. Ann. Stat. **33**, 1380–1403 (2005)
14. Ezzahrioui, M., Ould-Saïd, E.: Asymptotic results of a nonparametric conditional quantile estimator for functional time series. Commun. Stat. Theory Methods **37**, 2735–2759 (2008)
15. Gardes, L., Girard, S.: A moving window approach for nonparametric estimation of the conditional tail index. J. Multivar. Anal. **99**, 2368–2388 (2008)
16. Gardes, L., Girard, S.: Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels. Extremes **13**, 177–204 (2010)
17. Gardes, L., Girard, S.: Functional kernel estimators of large conditional quantiles. Electron. J. Stat. **6**, 1715–1744 (2012)
18. Gardes, L., Girard, S., Lekina, A.: Functional nonparametric estimation of conditional extreme quantiles. J. Multivar. Anal. **101**, 419–433 (2010)
19. Girard, S., Jacob, P.: Frontier estimation via kernel regression on high power-transformed data. J. Multivar. Anal. **99**, 403–420 (2008)
20. Girard, S., Menneteau, L.: Central limit theorems for smoothed extreme value estimates of Poisson point processes boundaries. J. Stat. Plan. Inference **135**, 433–460 (2005)
21. Hall, P., Tajvidi, N.: Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. Stat. Sci. **15**, 153–167 (2000)
22. Jurecková, J.: Remark on extreme regression quantile. Sankhya 69, 87–100 (2007)

23. Klass, M.: The Robbins-Siegmund series criterion for partial maxima. Ann. Probab. **13**, 1369–1370 (1985)
24. Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Ergebnisse der Mathematik. Springer, New York (1991)
25. Ould-Saïd, E., Yahia, D., Necir, A.: A strong uniform convergence rate of a kernel conditional quantile estimator under random left-truncation and dependent data. Electron. J. Stat. **3**, 426–445 (2009)
26. Park, B.U.: On nonparametric estimation of data edges. J. Korean Stat. Soc. **30**, 265–280 (2001)
27. Samanta, T.: Non-parametric estimation of conditional quantiles. Stat. Probab. Lett. **7**, 407–412 (1989)
28. Smith, R.L.: Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone (with discussion). Stat. Sci. **4**, 367–393 (1989)
29. Stone, C.J.: Consistent nonparametric regression (with discussion). Ann. Stat. **5**, 595–645 (1977)
30. Stute, W.: Conditional empirical processes. Ann. Stat. **14**, 638–647 (1986)
31. Tsay, R.S.: Analysis of Financial Time Series. Wiley, New York (2002)

Exact Quadratic Error of the Local Linear Regression Operator Estimator for Functional Covariates

Amina Naceri, Ali Laksaci, and Mustapha Rachdi

Abstract In this paper, it is studied the asymptotic behavior of the nonparametric local linear estimation of the regression operator when the covariates are curves. Under some general conditions we give the exact expression involved in the leading terms of the quadratic error of this estimator. The obtained results affirm the superiority of the local linear modeling over the kernel method, in functional statistics framework.

1 Introduction

This paper deals with the nonparametric regression operator estimation by the local linear modeling (cf. [1]). This subject is motivated by the fact that the local polynomial estimation method has various advantages over the kernel method (cf. [1, 5, 8] and references therein, for finite/infinite dimensional frameworks). Moreover, the classical Nadaraya–Watson kernel method can be viewed as a particular case of this procedure.

Notice that, the statistical analysis of infinite dimensional data (FDA) has become a major topic of research in the last decade, as evidenced by several special issues of various statistical journals dedicated to this topic (cf. for instance, [4, 10, 15, 17]). Furthermore, the nonparametric treatment of such data has also been widely developed in the last few years (cf. [11, 12] for recent advances and references). Concerning the local linear estimation technique in the functional setup, the first results were given by Baìllo and Grané [2]. They obtained the local linear

A. Naceri • A. Laksaci (✉)
Laboratoire de Statistique et Processus Stochastiques, Université Djillali Liabès, BP. 89, Sidi Bel-Abbès 22000, Algeria
e-mail: amina_nrc@outlook.fr; alilak@yahoo.fr

M. Rachdi
Laboratoire AGIM FRE 3405 CNRS, University of Grenoble-Alpes, 38400 Saint-Martin-d'Hères, France

University of P. Mendès France (Grenoble 2), UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France
e-mail: Mustapha.Rachdi@agim.eu; mustapha.rachdi@upmf-grenoble.fr

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_5

estimation of the regression function when the explanatory variable takes values in a Hilbert space. Barrientos et al. [1], meanwhile, established the almost-complete convergence (with rate) of the proposed estimator. We refer, also, to [1] for another alternative version of the functional local linear modeling. Elsewhere, El Methni and Rachdi [6] developed this smoothing local linear estimation of the regression operator for functional fixed design data in both Hilbert and semi-metric spaces. More recently, Demongeot et al. [5] considered the local polynomial modeling of the conditional density function when the explanatory variable is of functional kind. Among the lot of papers on the local linear modeling in the nonfunctional case, we refer to the papers by Chu and Marron [3], Fan [7] and Fan and Yao [9], among others.

The main aim of this paper is then to determine, under general conditions, the exact rates, in the mean squared error of the local linear estimation of the regression operator, in the i.i.d. case, as proposed by [1]. Specifically, we give the exact expression involved in the leading terms of the quadratic error. We point out that our study highlights the structural axis of this subject, namely the "dimensionality" of the model. Moreover, our results confirm the superiority of the local linear smoothing over the kernel method in the bias terms.

This paper is organized as follows. In Sect. 2, we present our model. Then, we give some notations, hypotheses, and the presentation of the main results in Sect. 3. The proofs of the results are religated to the last section.

2 Functional Local Linear Modeling: Estimation and Comments

Consider n pairs of independent random variables (X_i, Y_i) for $i = 1, \ldots, n$ that we assume drawn from the pair (X, Y) which is valued in $\mathscr{F} \times \mathbb{R}$, where $\mathscr{F}$ is a semi-metric space equipped with a semi-metric d. Our main purpose is to estimate the regression function $m(x) = \mathbb{E}[Y|X = x]$. For this purpose, it is well known that the main idea, in the local linear smoothing, is based on the fact that the function $m(x)$ admits a linear approximation in the neighborhood of the conditioning point. Recall that, in the non-functional case, this linear approximation is due to a Taylor expansion of $m(\cdot)$ but, in the functional setup, such approximation can be expressed, for any z in the neighborhood of x by:

$$m(z) = m(x) + b\beta(z, x) + o(\beta(z, x)). \tag{1}$$

Then, the quantities $a = m(x)$ and b are estimated by minimizing the following quantity:

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} (Y_i - a - b\beta(X_i, x))^2 \, K(h^{-1}\delta(x, X_i)),$$

where the function K is a kernel, $h = h_{K,n}$ is a sequence of positive real numbers, and $\delta(.,.)$ and $\beta(.,.)$ are two known bi-functionals defined from $\mathscr{F} \times \mathscr{F}$ into $\mathbb{R}$ such that:

$$\forall \xi \in \mathscr{F},\ \beta(\xi, \xi) = 0 \text{ and } d(.,.) = |\delta(.,.)|.$$

Such fast version of functional local linear estimation has been proposed by Barrientos et al. [1]. They showed, by a simple algebra, that $\hat{m}(x) = \hat{a}$ can be explicitly expressed by:

$$\hat{m}(x) = \frac{\sum_{i,j=1}^{n} W_{ij}(x) Y_j}{\sum_{i,j=1}^{n} W_{ij}(x)},$$

where

$$W_{ij}(x) = \beta(X_i, x)\left(\beta(X_i, x) - \beta(X_j, x)\right) K(h^{-1}\delta(x, X_i))K(h^{-1}\delta(x, X_j))$$

with the convention $0/0 = 0$.

Obviously, if $b = 0$, then we obtain the Nadaraya–Watson estimator studied, in the functional case, in [14] and the references therein. It is worth to noting that, under the fact that $\sum_{i,j=1}^{n} \beta_j W_{ij} = 0$, we can reformulate the bias term of the estimator $\hat{m}(x)$ as follows:

$$\begin{aligned}
\mathbb{E}\left[\hat{m}(x)\right] - m(x) &= \mathbb{E}\left[\frac{\sum_{i,j=1}^{n} W_{ij}(x)\left(Y_j - m(X_j)\right)}{\sum_{i,j=1}^{n} W_{ij}(x)}.\right] \\
&\quad + \mathbb{E}\left[\frac{\sum_{i,j=1}^{n} W_{ij}(x)\left(m(X_j) - m(x) + b\beta(X_j, x)\right)}{\sum_{i,j=1}^{n} W_{ij}(x)}\right] \\
&= \mathbb{E}\left[\frac{\sum_{i,j=1}^{n} W_{ij}(x)\left(Y_j - m(X_j)\right)}{\sum_{i,j=1}^{n} W_{ij}(x)}\right] \\
&\quad + \mathbb{E}\left[\frac{\sum_{i,j=1}^{n} W_{ij}(x)\left(o(\beta(X_j, x))\right)}{\sum_{i,j=1}^{n} W_{ij}(x)}\right].
\end{aligned}$$

Clearly, under the expansion (1), we obtain a bias term of order $o(h)$. Undoubtedly, this bias term is significantly better than in the kernel case, studied in [13], which is of order $O(h)$. Thus, we can say that the functional local linear modeling is not a simple generalization of the kernel method, rather more than that it is an alternative approach that has important advantages, in particular in the rate of convergence of the bias term (cf. Theorem 1 and Remark 2 below).

3 Main Results

In the remainder of this paper, we set:

$$\phi_x(r_1, r_2) = \mathbb{P}(r_2 \leq \delta(x, X) \leq r_1)$$

and

$$\Psi(s) = \mathbb{E}\left[m(X) - m(x) | \beta(x, X) = s\right].$$

We assume the following hypotheses:

(H1) For any $r > 0$, $\phi_x(r) := \phi_x(-r, r) > 0$ and there exists a function $\chi_x(\cdot)$ such that:

$$\forall t \in (-1, 1), \ \lim_{h \to 0} \frac{\phi_x(th, h)}{\phi_x(h)} = \chi_x(t).$$

(H2) The first (resp. the second) derivative of Ψ at 0 exists.

(H3) The bi-functional operator β is such that:

$$\text{for all } z \in \mathscr{F}, \ C_1 \left|\delta(x, z)\right| \leq \left|\beta(x, z)\right| \leq C_2 \left|\delta(x, z)\right|,$$

where $C_1 > 0, \ C_2 > 0$

$$\sup_{u \in B(x,r)} \left|\beta(u, x) - \delta(x, u)\right| = o(r)$$

and

$$h \int_{B(x,h)} \beta(u, x) dP_X(u) = o\left(\int_{B(x,h)} \beta^2(u, x)\, dP_X(u)\right),$$

where $B(x, r) = \{z \in \mathscr{F} \ : \ |\delta(x, z)| \leq r\}$ and $dP_X(x)$ is the probability distribution of X.

(H4) The kernel K is a positive, differentiable function which is supported within $(-1, 1)$. Moreover, its derivative K' satisfies $K'(t) < 0$, for $-1 \leq t < 1$ and $K(1) > 0$.

(H5) The function $m_2(\cdot) = \mathbb{E}[Y^2 | X = \cdot]$ is continuous in a neighborhood of x.

(H6) The bandwidth h satisfies :

$$\lim_{n \to \infty} h = 0, \ \text{ and } \lim_{n \to \infty} n\, \phi_x(h) = \infty.$$

Remark 1 Remark that assumptions (H1) and (H2) are an adaptations of conditions H_1 and H_3 in [13], when one replaces the semi-metric d by some bi-functional operator δ. The second part of the assumption (H3) has been introduced and commented in [1]. Readers would find, in this last, several examples of bi-functional operators δ and β which satisfy this condition. Finally, conditions (H4)–(H6) are classically used and are standard in the context of the quadratic error determination in functional statistics.

Theorem 1 *Under assumptions (H1)–(H6), we have that:*

$$\mathbb{E}\left[\hat{m}(x)-m(x)\right]^2 = B_K^2(x)h^4 + \frac{V_K(x)}{n\phi_x(h)} + o(h^4) + o\left(\frac{1}{n\ \phi_x(h)}\right), \quad (2)$$

where

$$B_K(x) = \frac{1}{2}\Psi''(0)\left[\frac{K(1)-\int_{-1}^{1}(u^2K(u))'\chi_x(u)du}{K(1)-\int_{-1}^{1}K'(u)\chi_x(u)du}\right]$$

and

$$V_K(x) = \left(m_2(x)-m^2(x)\right)\left[\frac{K^2(1)-\int_{-1}^{1}(K^2(u))'\chi(u)du}{\left(K(1)-\int_{-1}^{1}(K(u))'\chi(u)du\right)^2}\right].$$

Remark 2 We point out that Theorem 1 shows that the gain in the bias term is more important than $o(h)$ which is obtained under assumption (1) (cf. the discussion in Sect. 2). Undoubtedly, this gain is due to the additional regularity condition (H2). However, the variance term is exactly the same as for the classical kernel estimator.

Proof of Theorem 1 It is known that:

$$\mathbb{E}\left[\hat{m}(x)-m(x)\right]^2 = \left[E\left(\hat{m}(x)\right)-m(x)\right]^2 + \mathrm{Var}\left[\hat{m}(x)\right].$$

Then the proof of this Theorem is based on the separate calculations of the bias and the variance terms of the estimator $\hat{m}(x)$. For both quantities, we put:

$$\hat{g}(x) = \frac{1}{n\,(n-1)EW_{12}(x)}\sum_{i\neq j,1}^{n} W_{ij}(x)Y_j$$

and

$$\hat{f}(x) = \frac{1}{n\,(n-1)EW_{12}(x)}\sum_{i\neq j,1}^{n} W_{ij}(x).$$

So

$$\hat{m}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}.$$

Elsewhere, we have the following decomposition, for $z \neq 0$ and $p \in \mathbb{N}^*$:

$$\frac{1}{z} = 1 - (z-1) + \cdots + (-1)^p (z-1)^p + (-1)^{p+1} \frac{(z-1)^{p+1}}{z}.$$

Particularly, for $z = \hat{f}(x)$ and $p = 1$ we get:

$$\hat{m}(x) - m(x) = (\hat{g}(x) - m(x)) - (\hat{g}(x) - \mathbb{E}\hat{g}(x)) \left(\hat{f}(x) - 1\right) \\ -\mathbb{E}\hat{g}(x) \left(\hat{f}(x) - 1\right) + \left(\hat{f}(x) - 1\right)^2 \hat{m}(x).$$

Hence,

$$\mathbb{E}\left[\hat{m}(x)\right] - m(x) = (\mathbb{E}\hat{g}(x) - m(x)) - \mathrm{Cov}(\hat{g}(x), \hat{f}(x)) \\ +E\left(\hat{f}(x) - \mathbb{E}\hat{f}(x)\right)^2 \hat{m}(x).$$

Consequently, the bias term can be expressed by:

$$\mathbb{E}\left[\hat{m}(x)\right] - m(x) = (\mathbb{E}\hat{g}(x) - m(x)) - \mathrm{Cov}(\hat{g}(x), \hat{f}(x)) + O\left(\mathrm{Var}\left(\hat{f}(x)\right)\right).$$

Finally, the proof of Theorem 1 is based on Lemmas 1 and 2, below.

Lemma 1 *Under Assumptions (H1)–(H4) and (H6), we have:*

$$\mathbb{E}\left[\hat{g}(x)\right] - m(x) = B_K(x, y)h^2 + o(h^2) + O\left(\frac{1}{n\phi_x(h)}\right).$$

Lemma 2 *Under Assumptions (H1) and (H4)–(H6), we have:*

$$Var\left[\hat{m}(x)\right] = V_K(x) \left[\frac{K^2(1) - \int_{-1}^{1} (K^2(u))' \chi(u) du}{\left(K(1) - \int_{-1}^{1} (K(u))' \chi(u) du\right)^2}\right] + o\left(\frac{1}{n\phi_x(h)}\right).$$

Furthermore,

$$Cov(\hat{g}(x), \hat{f}(x)) = O\left(\frac{1}{n\phi_x(h)}\right)$$

and

$$Var\left[\hat{f}(x)\right] = O\left(\frac{1}{n\phi_x(h)}\right).$$

Appendix: Proofs

In what follows, when no confusion is possible, we will denote by C and C' some strictly positive generic constants. Moreover, we set for all $i, j = 1, \ldots, n$ and for a fixed $(x, y) \in \mathscr{F} \times \mathbb{R}$:

$$K_i = K(h^{-1}\delta(x, X_i)), \;\; \beta_i = \beta(X_i, x) \text{ and } W_{ij}(x) = W_{ij}.$$

Proof of Lemma 1 We have:

$$\begin{aligned}\mathbb{E}\left[\hat{g}(x)\right] &= \mathbb{E}\left[\frac{1}{n(n-1)\mathbb{E}[W_{12}]}\sum_{j\neq i,1}^{n} W_{ij}Y_j\right] \\ &= \frac{\mathbb{E}[W_{12}Y_2]}{\mathbb{E}[W_{12}]} = \frac{1}{\mathbb{E}[W_{12}]}\mathbb{E}\left[W_{12}\mathbb{E}[Y_2|X_2]\right]. \end{aligned} \tag{3}$$

Then, it follows from (3) and the definition of the operator m that:

$$\mathbb{E}\left[\hat{g}(x)\right] = \frac{1}{\mathbb{E}[W_{12}]}\mathbb{E}\left[W_{12}m(X_2)\right].$$

Now, by the same arguments as those used in [13], for the regression operator estimation, we show that:

$$\begin{aligned}\mathbb{E}\left[W_{12}m(X_2)\right] &= m(x)\mathbb{E}[W_{12}] + \mathbb{E}\left[W_{12}\left(m(X_2) - m(x)\right)\right] \\ &= m(x)\mathbb{E}[W_{12}] + \mathbb{E}\left[W_{12}\mathbb{E}\left[m(X_2) - m(x)|\beta(x, X_2)\right]\right] \\ &= m(x)\mathbb{E}[W_{12}] + \mathbb{E}\left[W_{12}\Psi\left(\beta(x, X_2)\right)\right]\end{aligned}$$

and since $\mathbb{E}\left[\beta(x, X_2)W_{12}\right] = 0$ and $\Psi(0) = 0$, we obtain:

$$\mathbb{E}\left[W_{12}\Psi\left(\beta(x, X_2)\right)\right] = \frac{1}{2}\Psi''(0)\mathbb{E}\left[\beta^2(x, X_2)W_{12}\right] + o\left(\mathbb{E}\left[\beta^2(x, X_2)W_{12}\right]\right).$$

Then:

$$\mathbb{E}\left[\hat{g}(x)\right] = m(x) + \Psi_0''(0)\frac{\mathbb{E}\left[\beta^2(x, X_2)W_{12}\right]}{2\mathbb{E}[W_{12}]} + o\left(\frac{\mathbb{E}\left[\beta^2(x, X_2)W_{12}\right]}{\mathbb{E}[W_{12}]}\right).$$

Moreover, it is clear that:

$$\begin{aligned}\mathbb{E}\left[\beta(x, X_2)^2W_{12}\right] &= \left(\mathbb{E}\left[K_1\beta_1^2\right]\right)^2 - \mathbb{E}[K_1\beta_1]\mathbb{E}[K_1\beta_1^3] \\ \mathbb{E}\left[W_{12}\right] &= \mathbb{E}[K_1\beta_1^2]EK_1 - (\mathbb{E}[K_1\beta_1])^2\end{aligned}$$

and, under the Assumption (H4), we obtain that:

$$\text{for all } a > 0,\ \mathbb{E}[K_1^a \beta_1] \leq C \int_{B(x,h)} \beta(u,x) dP_X(u).$$

So, by using the last part of the Assumption (H3), we get:

$$h\mathbb{E}[K_1^a \beta_1] = o\left(\int_{B(x,h)} \beta^2(u,x) dP_X(u) \right) = o(h^2 \phi_x(h))$$

which allows to write:

$$\mathbb{E}[K_1^a \beta_1] = o(h\phi_x(h)). \tag{4}$$

Moreover, for all $b > 1$, we can write:

$$\mathbb{E}[K_1^a \beta_1^b] = \mathbb{E}[K_1^a \delta^b(x, X_1)] + \mathbb{E}\left[K_1(\beta^b(X_1, x) - \delta^b(x, X_1)) \right].$$

Then, the second part of the Assumption (H3) implies that:

$$\begin{aligned}
&\mathbb{E}\left[K_1^a(\beta^b(X_1, x) - \delta^b(x, X_1)) \right] \\
&= \mathbb{E}\left[K_1^a I_{B(x,h)} (\beta(X_1, x) - \delta(x, X_1)) \sum_{l=0}^{b-1} (\beta(X_1, x))^{b-1-l} (\delta(x, X_1))^l \right] \\
&\leq \sup_{u \in B(x,h)} |\beta(u,x) - \delta(x,u)| \sum_{l=0}^{b-1} \mathbb{E}\left[K_1^a I_{B(x,h)} |\beta(X_1, x)|^{b-1-l} |\delta(x, X_1)|^l) \right],
\end{aligned}$$

whereas the first part of the Assumption (H3) gives:

$$I_{B(x,h)} |\beta(X_1, x)| \leq I_{B(x,h)} |\delta(x, X_1)|.$$

Thus, it follows:

$$\begin{aligned}
\mathbb{E}\left[K_1^a(\beta^b(X_1, x) - \delta^b(x, X_1)) \right] &\leq b \sup_{u \in B(x,h)} |\beta(u,x) - \delta(x,u)| |\mathbb{E}[K_1^a |\delta|^{b-1}(x, X_1)] \\
&\leq b \sup_{u \in B(x,h)} |\beta(u,x) - \delta(x,u)| h^{b-1} \mathbb{E}[K_1^a] \\
&\leq b \sup_{u \in B(x,h)} |\beta(u,x) - \delta(x,u)| h^{b-1} \phi_x(h)
\end{aligned}$$

which allows to write:

$$\mathbb{E}[K_1^a \beta_1^b] = \mathbb{E}[K_1^a \delta^b(x, X_1)] + o(h^b \phi_x(h)).$$

Concerning the term $\mathbb{E}[K_1^a\delta^b]$, we write:

$$
\begin{aligned}
h^{-b}\mathbb{E}[K_1^a\delta^b] &= \int v^b K^a(v)dP_X^{h^{-1}\delta(x,X_1)}(v)\\
&= \int_{-1}^{1}\left[K^a(1)-\int_v^1 (u^bK^a(u))'du\right]dP_X^{h^{-1}\delta(x,X_1)}(v)\\
&= K^a(1)\phi_x(h)-\int_{-1}^{1}(u^bK^a(u))'\phi_x(uh,h)du\\
&= \phi_x(h)\left(K^a(1)-\int_{-1}^{1}(u^bK^a(u))'\frac{\phi_x(uh,h)}{\phi_x(h)}du\right).
\end{aligned}
$$

Finally, under the Assumption (H1), we get:

$$
\mathbb{E}[K_1^a\beta_1^b] = h^b\phi_x(h)\left(K^a(1)-\int_{-1}^{1}(u^bK^a(u))'\chi_x(u)du\right)+o(h^b\phi_x(h)). \tag{5}
$$

It follows that:

$$
\frac{\mathbb{E}\left[\beta^2(x,X_2)W_{12}\right]}{\mathbb{E}[W_{12}]} = h^2\left[\frac{K(1)-\int_{-1}^{1}(u^2K(u))'\chi_x(u)du}{K(1)-\int_{-1}^{1}K'(u)\chi_x(u)du}\right]+o(h^2).
$$

Consequently:

$$
\mathbb{E}\left[\hat{g}(x)\right] = m(x)+\frac{h^2}{2}\Psi_0''(0)\left[\frac{K(1)-\int_{-1}^{1}(u^2K(u))'\chi_x(u)du}{K(1)-\int_{-1}^{1}K'(u)\chi_x(u)du}\right]+o(h^2).
$$

■

Proof of Lemma 2 For this Lemma, we use the same ideas of Sarda and Vieu [16] to show that

$$
\begin{aligned}
\mathrm{Var}\left[\hat{m}(x)\right] = \mathrm{Var}\left[\hat{g}(x)\right] &- 2(\mathbb{E}\hat{g}(x))\mathrm{Cov}(\hat{g}(x),\hat{f}(x))\\
&+(\mathbb{E}\hat{g}(x))^2\mathrm{Var}(\hat{f}(x))+o\left(\frac{1}{n\phi_x(h)}\right).
\end{aligned}
$$

It is clear that:

$$\begin{aligned}
\mathrm{Var}\,(\hat{g}(x)) &= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2}\mathrm{Var}\left(\sum_{i\neq j=1}^{n} W_{ij}Y_j\right) \\
&= \frac{1}{(n(n-1)(EW_{12}))^2}\Big(n(n-1)\mathbb{E}[W_{12}^2Y_2^2] + n(n-1)\mathbb{E}[W_{12}W_{21}Y_2Y_1] \\
&\quad +n(n-1)(n-2)\mathbb{E}[W_{12}W_{13}Y_2Y_3] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{23}Y_2Y_3] \\
&\quad +n(n-1)(n-2)\mathbb{E}[W_{12}W_{31}Y_2Y_1] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{32}Y_2^2] \\
&\quad -n(n-1)(4n-6)(\mathbb{E}[W_{12}Y_2])^2\Big)\,.
\end{aligned}$$

Observe that the terms of the first line are negligible compared to other terms which are multiplied by $n(n-1)(n-2)$. Furthermore,

$$\begin{aligned}
\mathbb{E}[W_{12}^2Y_2^2] &= O(h^4\phi_x^2(h)), \\
\mathbb{E}[W_{12}W_{21}Y_1Y_2] &= O(h^4\phi_x^2(h)), \\
\mathbb{E}[W_{12}W_{13}Y_2Y_3] &= (m(x))^2\mathbb{E}[\beta_1^4K_1^2](\mathbb{E}[K_1])^2 + o(h^4\phi_x^3(h)), \\
\mathbb{E}[W_{12}W_{23}Y_2Y_3] &= (m(x))^2\mathbb{E}[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1]) + o(h^4\phi_x^3(h)), \\
\mathbb{E}[W_{12}W_{31}Y_2Y_1] &= (m(x))^2\mathbb{E}[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1]) + o(h^4\phi_x^3(h)), \\
\mathbb{E}[W_{12}W_{32}Y_2^2] &= (m_2(x))(\mathbb{E}[\beta_1^2K_1])^2(\mathbb{E}[K_1^2]) + o(h^4\phi_x^3(h)), \\
\mathbb{E}[W_{12}Y_2] &= O(h^2\phi_x^2(h)).
\end{aligned}$$

Therefore, the leading term in the expression of $\mathrm{Var}\,(\hat{g}(x))$ is:

$$\begin{aligned}
&\frac{n(n-1)(n-2)}{(n(n-1)\mathbb{E}[W_{12}])^2}\Big((m(x))^2\left(\mathbb{E}[\beta_1^4K_1^2](\mathbb{E}[K_1])^2 + 2BBe[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1])\right) \\
&\quad +(m_2(x))(\mathbb{E}[\beta_1^2K_1])^2(\mathbb{E}[K_1^2]) + o(h^4\phi_x^3(h))\Big)\,.
\end{aligned}$$

Concerning the covariance term, we have by the same fashion:

$$\begin{aligned}
\mathrm{Cov}(\hat{g}(x),\hat{f}(x)) &= \frac{1}{(n(n-1)\mathbb{E}[W_{12}])^2}\mathrm{Cov}\left(\sum_{\substack{i,j=1\\ i\neq j}}^{n} W_{ij}Y_j, \sum_{\substack{i',j'=1\\ i'\neq j'}}^{n} W_{i'j'}\right) \\
&= \frac{1}{(n(n-1)EW_{12}))^2}\Big[n(n-1)\mathbb{E}[W_{12}^2Y_2] + n(n-1)\mathbb{E}[W_{12}W_{21}Y_2] \\
&\quad +n(n-1)(n-2)\mathbb{E}[W_{12}W_{13}Y_2] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{23}Y_2] \\
&\quad +n(n-1)(n-2)\mathbb{E}[W_{12}W_{31}Y_2] + n(n-1)(n-2)\mathbb{E}[W_{12}W_{32}Y_2] \\
&\quad -n(n-1)(4n-6)(\mathbb{E}[W_{12}Y_2]\mathbb{E}[W_{12}])\Big]
\end{aligned}$$

with

$$
\begin{aligned}
\mathbb{E}[W_{12}^2 Y_2] &= O(h^4\phi_x^2(h)),\\
\mathbb{E}[W_{12}W_{21}Y_2] &= O(h^4\phi_x^2(h)),\\
\mathbb{E}[W_{12}W_{13}Y_2] &= (m(x))\mathbb{E}[\beta_1^4K_1^2](\mathbb{E}[K_1])^2 + o(h^4\phi_x^3(h)),\\
\mathbb{E}[W_{12}W_{23}Y_2] &= (m(x))\mathbb{E}[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[W_{12}W_{31}Y_2] &= (m(x))\mathbb{E}[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[W_{12}W_{32}Y_2] &= (m(x))(\mathbb{E}[\beta_1^2K_1])^2(\mathbb{E}[K_1^2]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[W_{12}Y_1] &= O(h^2\phi_x^2(h)).
\end{aligned}
$$

Therefore, the leading term in the expression of $\mathrm{Cov}(\hat{g}(x),\hat{f}(x))$ is:

$$
\begin{aligned}
&\frac{n(n-1)(n-2)}{(n(n-1)\mathbb{E}[W_{12}])^2}\left(m(x)\left(\mathbb{E}[\beta_1^4K_1^2](\mathbb{E}[K_1])^2 + 2\mathbb{E}[\beta_1^2K_1](\mathbb{E}[\beta_1^2K_1^2]\mathbb{E}[K_1])\right.\right.\\
&\qquad \left.\left.+(\mathbb{E}[\beta_1^2K_1])^2(\mathbb{E}[K_1^2])\right) + o(h^4\phi_x^3(h))\right).
\end{aligned}
$$

Finally, for $\mathrm{Var}\left(\hat{f}(x)\right)$

$$
\begin{aligned}
\mathrm{Var}\left(\hat{f}(x)\right) = \frac{1}{(n(n-1)(\mathbb{E}W_{12}))^2}&\Big[n(n-1)\mathbb{E}[[W_{12}^2] + n(n-1)\mathbb{E}[[W_{12}W_{21}]\\
&+n(n-1)(n-2)\mathbb{E}[[W_{12}W_{13}] + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{23}]\\
&+n(n-1)(n-2)\mathbb{E}[[W_{12}W_{31}] + n(n-1)(n-2)\mathbb{E}[[W_{12}W_{32}]\\
&-n(n-1)(4n-6)(\mathbb{E}[[W_{12}])^2\Big]
\end{aligned}
$$

and similarly to the previous cases:

$$
\begin{aligned}
\mathbb{E}[W_{12}^2] &= O(h^4\phi_x^2(h)),\\
\mathbb{E}[[W_{12}W_{21}] &= O(h^4\phi_x^2(h)),\\
\mathbb{E}[[W_{12}W_{13}] &= \mathbb{E}[[\beta_1^4K_1^2](\mathbb{E}[[K_1])^2 + o(h^4\phi_x^3(h)),\\
\mathbb{E}[[W_{12}W_{23}] &= \mathbb{E}[[\beta_1^2K_1](\mathbb{E}[[\beta_1^2K_1^2]\mathbb{E}[[K_1]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[[W_{12}W_{31}] &= \mathbb{E}[[\beta_1^2K_1](\mathbb{E}[[\beta_1^2K_1^2]\mathbb{E}[[K_1]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[[W_{12}W_{32}] &= (\mathbb{E}[[\beta_1^2K_1])^2(\mathbb{E}[[K_1^2]) + o(h^4\phi_x^3(h)),\\
\mathbb{E}[[W_{12}] &= O(h^2\phi_x^2(h)).
\end{aligned}
$$

Therefore,

$$\mathrm{Var}\,(\hat{m}(x)) = \frac{(m_2(x) - m^2(x))}{n\phi_x(h)} \left[\frac{\left(K^2(1) - \int_{-1}^{1} (K^2(u))' \chi(u)du\right)}{\left(K(1) - \int_{-1}^{1} (K(u))' \chi(u)du\right)^2} \right] + o\left(\frac{1}{n\phi_x(h)}\right).$$

which completes the proof. ■

Acknowledgements We are indebted to Campus France CMCU for supporting us with the grant PHC Maghreb SCIM.

References

1. Barrientos-Marin, J., Ferraty, F., Vieu, P.: Locally modelled regression and functional data. J. Nonparametr. Stat. **22**, 617–632 (2010)
2. Baìllo, A., Grané, A.: Local linear regression for functional predictor and scalar response. J. Multivar. Anal. **100**, 102–111 (2009)
3. Chu, C.-K., Marron, J.-S.: Choosing a kernel regression estimator. Stat. Sci. **6**, 404–436 (1991)
4. Davidian, M., Lin, X., Wang, J.L.: Introduction. Stat. Sinica **14**, 613–614 (2004)
5. Demongeot, J., Laksaci, A., Madani, F., Rachdi, M.: Functional data: local linear estimation of the conditional density and its application. Statistics **47**, 26–44 (2013)
6. El Methni, M., Rachdi, M.: Local weighted average estimation of the regression operator for functional data. Commun. Stat. Theory Methods **40**, 3141–3153 (2010)
7. Fan, J.: Design-adaptive nonparametric regression. J. Am. Stat. Assoc. **87**, 998–1004 (1992)
8. Fan, J., Gijbels, I.: Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability, vol. 66. Chapman & Hall, London (1996)
9. Fan, J., Yao, Q.: Nonlinear Time Series: Nonparametric and Parametric Methods. Springer, New York (2003)
10. Ferraty, F.: High-dimensional data: a fascinating statistical challenge. J. Multivar. Anal. **101**, 305–30 (2010)
11. Ferraty, F., Romain, Y.: The Oxford Handbook of Functional Data Analysis. Oxford University Press, Oxford (2011)
12. Ferraty, F., Vieu, P.: Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer, New York (2006)
13. Ferraty, F., Mas, A., Vieu, P.: Nonparametric regression on functional data: inference and practical aspects. Aust. N. Z. J. Stat. **49**, 267–286 (2007)
14. Ferraty, F., Laksaci, A., Tadj, A., Vieu, P.: Rate of uniform consistency for nonparametric estimates with functional variables. J. Stat. Plan. Inference **140**, 335–352 (2010)
15. Gonzalez Manteiga, W., Vieu, P.: Statistics for functional data. Comput. Stat. Data Anal. **51**, 4788–4792 (2007
16. Sarda, P., Vieu, P.: Kernel regression. In: Schimek, M.G. (ed.) Smoothing and Regression: Approaches, Computation and Application. Wiley Series in Probability and Statistics, pp. 43–70. Wiley, Chichester, New York (2000)
17. Valderrama, M.J.: An overview to modelling functional data. Comput. Stat. **22**, 331–334 (2000)

Estimation of a Loss Function for Spherically Symmetric Distribution with Constraints on the Norm

Idir Ouassou

Abstract In this paper, we consider the problem of estimating the quadratic loss of point estimators of a location parameter θ, for a family of symmetric distribution with known scale parameter, subject to different constraints to be satisfied by the norm of θ and when a residual vector U is available. We compare the robust and non-robust estimators. Moreover we give sufficient conditions on the distribution for the domination of competing estimators. In particular this result remains true for t-distributions when the dimension of the residual vector is sufficiently large. The upper and lower bounds on the risk are exact at $\theta = 0$.

1 Introduction

Consider the problem of estimating the loss incurred using least squares estimators of a location parameter of a spherically symmetric distribution when the scale parameter is known, its norm satisfies different constraints and the residual vector U is available. This problem has been firstly considered by Lehmann and Scheffé [6] who estimated the power of a statistic test. Later on Fourdrinier and Wells [3] by Johnstone [5] Lele [7], Ouassou and Rachdi [4, 8, 9] have studied this problem in a variety of situation.

The present work study a class of estimators which improve an unbiased loss estimator $\lambda_u(x) = p\|U\|^2/k$ of the usual minimax estimator X. A particular important class of such estimators is the class $\lambda^c(X) = \lambda_u - c/\|X\|^2$. An alternative class, when a residual vector U is available, is the class of robust estimators $\lambda^c(X, U) = \lambda_u - c\|U\|^2/\|X\|^2$.

The paper is organized as follows. In Sect. 2, we give an expression of the risk of the estimator λ^c and of the robust estimators λ^c_R. Section 3 is devoted to upper and lower bounds on expectations $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^2\right]$ and $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$. Then we derive an upper and a lower bound on the risk on non-robust and robust estimators.

I. Ouassou (✉)
National School of Applied Sciences, Cadi Ayyad University, Av. Abdelkrim Khattabi, BP. 575, Marrakesh, Morocco
e-mail: i.ouassou@uca.ma

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_6

These bounds are similar in spirit to those developed by Casella and Hwang [1] and by Fourdrinier and Strawderman [2] for the estimation of the location parameter. Section 4 is devoted to an application of the results of the previous section which gives sufficient conditions for the domination of the robust estimator over the non-robust one.

We also give sufficient conditions for domination when θ is in a neighborhood of 0 and in a neighborhood of infinity.

Section 5 gives various examples of distributions illustrating the phenomenon. A basic example is the t-distribution when the dimension of the residual vector used for estimating the variance is large enough.

Finally, we present an appendix which contains technical lemmas used to prove various results and establish important estimates of the risk bounds.

2 Robust and Non-robust Estimators

Given a random vector $(X, U) \in \mathbf{R}^{p+k}$ which has a spherically symmetric distribution with location parameter $(\theta, 0) \in \mathbf{R}^{p+k}$ such that $\|\theta\| = \eta$ for a known η. The dimension of the sub-vectors X and θ is p while that of U and 0 is k. We are interested in the estimation of the quadratic loss function $\|X-\theta\|^2$ of the usual least square estimator X of θ. The estimator X is the orthogonal projector from $\mathbf{R}^{p+k}$ onto $\mathbf{R}^p$. An unbiased estimator of $\|X-\theta\|^2$, given by Fourdrinier and Wells [3], is

$$\lambda_u = \frac{p}{k}\|U\|^2,$$

a competitive class of estimators are of the form

$$\lambda^c(X) = \lambda_u - c\frac{1}{\|X\|^2} \tag{1}$$

where c is a positive constant. Others competing estimators, improving λ_u, are those which take into account the residual vector U, that is,

$$\lambda_R^c(X, U) = \lambda_u - c\frac{\|U\|^4}{\|X\|^2}. \tag{2}$$

We will refer to the estimators of the form (2) as robust estimators, the robust character here refers to the class of all spherical laws and the presence of U.

For any θ, we denote by $\mathbb{E}_\theta$ the expectation with respect to the underlying spherically symmetric distribution with mean θ. Hence the risk of any estimator λ is given by

$$R(\lambda, X, \theta) = \mathbb{E}_\theta\left[(\lambda - \|X-\theta\|^2)^2\right].$$

The risk of the estimators given in (1) and (2) can be calculated simultaneously by the risk of the following estimator

$$\lambda_\alpha^c(X, U) = \lambda_u - c\frac{\|U\|^\alpha}{\|X\|^2}. \tag{3}$$

where $\alpha \in \mathbf{N}$.

The finiteness of such risks is guaranteed as soon as the second moment of the spherical distribution exists and $\mathbb{E}_\theta\left[\|U\|^{2\alpha}/\|X\|^4\right] < \infty$ and $\mathbb{E}_\theta\left[\|U\|^{\alpha+2}/\|X\|^2\right] < \infty$.

The expression of the risk of λ_α^c is given by the next proposition.

Proposition 2.1 *For any $\theta \in \mathbb{R}^p$, the risk $R(\lambda_\alpha^c, X, \theta)$ of λ_α^c equals to*

$$R(\lambda_u, X, \theta) + c^2\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right] - 2c\frac{p-4}{(k+\alpha)(k+\alpha+2)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right]$$
$$-2c\frac{p\alpha}{k(k+\alpha)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]. \tag{4}$$

Proof of the Proposition 2.1 Let θ be fixed in $\mathbb{R}^p$, we can write

$$\begin{aligned} R(\lambda_c^\alpha, \lambda_u, \theta) &= \mathbb{E}_\theta\left[\left(\lambda_c^\alpha - \|X-\theta\|^2\right)^2\right] \\ &= R(\lambda_u, X, \theta) + c^2\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right] - 2c\frac{p}{k}\mathbb{E}_\theta\left[\frac{\|U\|^{2+\alpha}}{\|X\|^2}\right] \\ &\quad +2c\mathbb{E}_\theta\left[\|X-\theta\|^2\frac{\|U\|^\alpha}{\|X\|^2}\right]. \end{aligned}$$

According to Lemma A.1 in [3] applied with $g(x) = 1/||x||^2$, we have

$$\begin{aligned} &(k+\alpha)\mathbb{E}_{R,\theta}\left[\|X-\theta\|^2\frac{\|U\|^\alpha}{\|X\|^2}\right] \\ &= p\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right] + \frac{1}{k+\alpha+2}\mathbb{E}_{R,\theta}\left[\|U\|^{\alpha+4}\Delta\left(\frac{1}{\|X\|^2}\right)\right] \\ &= p\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right] - 2c\frac{p-4}{k+\alpha+2}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right]. \end{aligned}$$

Recalling that $\Delta\left(1/\|X\|^2\right) = -(p-4)/\|X\|^4$. Therefore the above risk expression is proved. ■

Remarks

1. As the risk $R(\lambda_\alpha^c, X, \theta)$ has a quadratic form then it is easy to deduce from the risk of λ_α^c, that, for any $\theta \in \mathbb{R}^p$, the constant $c_\alpha^\star$ for which the risk is minimum is

$$c_\alpha^\star(\eta) = \frac{\frac{p-4}{(k+\alpha)(k+\alpha+2)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right] + \frac{p\alpha}{k(k+\alpha)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]}{\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]} \tag{5}$$

and the corresponding risk is

$$R(\lambda_u, X, \theta) - \frac{\left(\frac{p-4}{(k+\alpha)(k+\alpha+2)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right] + \frac{p\alpha}{k(k+\alpha)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]\right)^2}{\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]}. \tag{6}$$

2. It is worth noting that, for the non-robust estimator ($\alpha = 0$), the optimal constant typically depends on θ and equals to

$$c_0^\star(\eta) = \frac{p-4}{k(k+2)} \frac{\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right]}{\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]} \tag{7}$$

and its risk equals to

$$R(\lambda_u, X, \theta) - \frac{\left(\frac{p-4}{k(k+2)}\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right]\right)^2}{\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]}. \tag{8}$$

3. For the robust estimator ($\alpha = 4$), the optimal constant depends on θ and is equal to

$$c(\eta) = \frac{p-4}{(k+4)(k+6)} + \frac{4p}{k(k+4)} \frac{\mathbb{E}_\theta\left[\frac{\|U\|^6}{\|X\|^2}\right]}{\mathbb{E}_\theta\left[\frac{\|U\|^8}{\|X\|^4}\right]} \tag{9}$$

and the corresponding risk is

$$R(\lambda_u, X, \theta) - \frac{\left(\frac{p-4}{(k+4)(k+6)}\mathbb{E}_\theta\left[\frac{\|U\|^8}{\|X\|^4}\right] + \frac{4p}{k(k+4)}\mathbb{E}_\theta\left[\frac{\|U\|^6}{\|X\|^2}\right]\right)^2}{\mathbb{E}_\theta\left[\frac{\|U\|^8}{\|X\|^4}\right]}. \tag{10}$$

4. However in the normal case $N(\theta, \sigma I_{p+k})$, the optimal constant $c^\star(\eta)$ for non-robust case does not depend on $\eta = \|\theta\|$ and is equal to $(p-4)\sigma^2$. In general, for the independence of $c^\star(\eta)$ on η, it would be sufficient that $\|U\|^2$ and $1/\|X\|^2$ are

uncorrelated for all θ. While, for the robust estimators λ_4^c the optimal constant $c_4^\star$ depends also on $\eta = \|\theta\|$ in the normal case. Indeed, a classical calculation based on the non-centered χ^2 distribution and the Poisson one leads to

$$c_4^\star(\eta) = \frac{p-4}{(k+4)(k+6)} - \frac{4p}{k(k+4)(k+6)} \frac{E^L\left[\frac{1}{p+2L-2}\right]}{E^L\left[\frac{1}{(p+2L-2)(p+2L-4)}\right]}$$

where L is a Poisson random variable with parameter $\eta/2$.

3 Bounds for $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^2\right]$ and $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$

The Proposition 1 indicates that the bounds of the risks of the estimators in (1) and (2) are of the form $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^2\right]$ and $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$, where q is an integer.

The following propositions yield such upper and lower bounds which are expressed, for any fixed $R \geq 0$ conditionally on the radius $R = (\|X-\theta\|^2 + \|U\|^2)^{1/2}$. Let us denote by $\mathbb{E}_{R,\theta}$ the expectation with respect to the uniform distribution $U_{R,\theta}$ on the sphere $S_{R,\theta} = \{y \in \mathbb{R}^{p+k}/ \ \|y-\theta\| = R\}$. Thus we can write $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right] = \mathbb{E}\left[\mathbb{E}_{R,\theta}\left(\|U\|^{2q}/\|X\|^4\right)/R\right]$, where E denotes the expectation with respect to the radial distribution.

In this section we give only the bounds of the expectation $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$ since the bounds of the second expectation are given in [2].

First we give an expression of $\mathbb{E}_{R,\theta}\left[\|U\|^{2q}/\|X\|^4\right]$ in terms of integrals with respect to a Beta distribution. For notational convenience we often use $B(\alpha,\beta,dv)$ for the density of the Beta distribution with parameters $\alpha > 0$ and $\beta > 0$.

Proposition 3.1 *For $p \geq 3$, any $R > 0$, any $\theta \in \mathbb{R}^p$ and any integer q such that $-\frac{k}{2} < q$, the expectation of $\|U\|^{2q}/\|X\|^4$ conditionally on the radius R is equal to*

$$\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] = \frac{\Gamma\left(\frac{k+p}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{k+p}{2}+q\right)\Gamma\left(\frac{k}{2}\right)} R^{2q}$$
$$\times \int_0^1\int_0^1 \frac{(R^2u+\|\theta\|^2)^2 + 4R^2\|\theta\|^2uv}{\left((R^2u+\|\theta\|^2)^2 - 4R^2\|\theta\|^2uv\right)^2} B\left(\frac{1}{2},\frac{p-1}{2},dv\right) B\left(\frac{p}{2},\frac{k}{2}+q,du\right).$$

At $\theta = 0$ this proposition, simplifies greatly in the following corollary.

Corollary 3.1 *Under the condition of Proposition 3.1, we have for $\theta = 0$ and $p > 4$*

$$\mathbb{E}_0\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] = \frac{B\left(\frac{k}{2}+q,\frac{p-4}{2}\right)}{B\left(\frac{k}{2},\frac{p}{2}\right)}\mathbb{E}\left[R^{2(q-2)}\right]$$

and in the particular case where $q = 0, 2, 4$, *we have*

$$\mathbb{E}_0\left[\frac{1}{\|X\|^4}\right] = \frac{(p+k-2)(p+k-4)}{(p-2)(p-4)}\mathbb{E}[R^4]$$
$$\mathbb{E}_0\left[\frac{\|U\|^4}{\|X\|^4}\right] = \frac{k(k+2)}{(p-2)(p-4)}$$
$$\mathbb{E}_0\left[\frac{\|U\|^8}{\|X\|^4}\right] = \frac{k(k+2)(k+4)(k+6)}{(p+k)(p+k+2)(p-2)(p-4)}\mathbb{E}[R^4].$$

The proposition below gives an upper bound of $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$.

Proposition 3.2 *Let q be an integer such that* $\frac{-k}{2} < q$. *If* $p \geq 8$ *then for any* $\theta \in \mathbb{R}^p$,

$$\mathbb{E}_\theta\left[\frac{||u||^{2q}}{||X||^4}\right] \leq \frac{\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{p+k}{2}+q\right)\Gamma\left(\frac{k}{2}\right)}\frac{(p+k+2q-2)(p+k+2q-4)}{(p-2)(p-4)} \tag{11}$$
$$\times\mathbb{E}\left[\frac{R^{2q}}{\left(R^2+\frac{p+k+2q-2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k+2q-4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right].$$

Corollary 3.2 *If* $p \geq 14$, *then, for* $q = 0,\ 2,\ 4$ *and for any* $\theta \in \mathbb{R}^p$, *we have*

$$\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right] \leq \frac{(p+k-2)(p+k-4)}{(p-2)(p-4)}$$
$$\times\mathbb{E}\left[\frac{1}{\left(R^2+\frac{p+k-2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k-4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right]$$
$$\leq \frac{k(k+2)}{(p-2)(p-4)}\mathbb{E}\left[\frac{R^4}{\left(R^2+\frac{p+k+2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]$$
$$\mathbb{E}_\theta\left[\frac{\|U\|^8}{\|X\|^4}\right]$$
$$\leq \frac{k(k+2)(k+4)(k+6)}{(p-4)(p-2)(p+k)(p+k+2)}$$
$$\times\mathbb{E}\left[\frac{R^8}{\left(R^2+\frac{p+k+6}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k+4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right].$$

The next proposition gives a lower bound of $\mathbb{E}_\theta\left[\|U\|^{2q}/\|X\|^4\right]$.

Proposition 3.3 *Let q be an integer such that $\frac{-k}{2} < q$. If $p \geq 5$, then for any $\theta \in \mathbb{R}^p$,*

$$\mathbb{E}_\theta\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] \geq \frac{\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{p+k}{2}+q\right)}\frac{(p+k+2q-2)(p+k+2q-4)}{(p-2)(p-4)}$$

$$\times\mathbb{E}\left[\frac{R^{2q}}{\left(R^2+\frac{p+K+2q-4}{p-4}\|\theta\|^2\right)^2}\right]. \tag{12}$$

Corollary 3.3 *If $p \geq 5$, then, for $q = 0,\ 2,\ 4$ and for any $\theta \in \mathbb{R}^p$, we have*

$$\begin{aligned}
\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]
&\geq \left(\frac{p+k-4}{p-4}\right)^2\mathbb{E}\left[\frac{1}{\left(R^2+\frac{p+k-4}{p-4}\|\theta\|^2\right)^2}\right]\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right]\\
&\geq \frac{p+k}{(p-4)^2}\frac{k(k+2)}{p+k+2}\mathbb{E}\left[\frac{R^4}{\left(R^2+\frac{p+k}{p-4}\|\theta\|^2\right)^2}\right]\mathbb{E}_\theta\left[\frac{\|U\|^8}{\|X\|^4}\right]\\
&\geq \frac{k(k+2)(k+4)(k+6)(p+k+4)}{(p-2)(p-4)^2(p+k)(p+k+2)(p+k+6)}\\
&\quad \mathbb{E}\left[\frac{R^8}{\left[R^2+\frac{p+K+4}{p-4}\|\theta\|^2\right]^2}\right].
\end{aligned}$$

3.1 Bounds for the Risk of $\boldsymbol{\lambda_R^c}$ and $\boldsymbol{\lambda^c}$

In this subsection, we give lower and upper bounds for the risk of the estimators λ^c (1) and the robust estimator λ_R^c (2). Their risks correspond respectively to the case $\alpha = 0$ and $\alpha = 4$ in proposition 2.1. See [1] and [2] for similar bounds in the estimation of the mean θ. Thus we have

$$R(\lambda^c, X, \theta) = R(\lambda_u, X, \theta) + c^2\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right] - 2c\frac{p-4}{k(k+2)}\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right] \tag{13}$$

and

$$R(\lambda_R^c, X, \theta) = R(\lambda_u, X, \theta) + c^2 \mathbb{E}_\theta \left[\frac{\|U\|^8}{\|X\|^4} \right] - 2c \frac{p-4}{(k+4)(k+6)} \mathbb{E}_\theta \left[\frac{\|U\|^8}{\|X\|^4} \right]$$
$$-8c \frac{p}{k+4} \mathbb{E}_\theta \left[\frac{\|U\|^6}{\|X\|^2} \right]. \tag{14}$$

In this context, bounds for $R(\lambda^c, X, \theta)$ and $R(\lambda_R^c, X, \theta)$ are immediately deduced from Corollaries 3.1 and 3.3.

In the next proposition, we give a lower and upper bounds of the risk difference between λ_c and $\lambda_R^{c_4}$ relative to the risk of the unbiased estimator λ_u. Denote by $\Delta\mathscr{R}_R(\theta) = R(\lambda_R^{c_4}, X, \theta) - R(\lambda_u, X, \theta)$ and $\Delta\mathscr{R}(\theta) = R(\lambda_c, X, \theta) - R(\lambda_u, X, \theta)$ these differences.

Proposition 3.4 *For $p \geq 14$ and $\theta \in \mathbb{R}^p$, the upper and lower bounds of $\Delta\mathscr{R}(\theta)$ are respectively*

$$c^2 \mathscr{C}_1 \mathbb{E}\left[\frac{1}{\left(R^2 + \frac{p+k-4}{p-4}\eta^2\right)^2} \right] - \frac{2c}{p-2} \mathbb{E}\left[\frac{R^4}{\left(R^2 + \frac{p+k+2}{p-2}\eta^2\right)\left(R^2 + \frac{(p+k)(p-14)}{(p-2)(p-4)}\eta^2\right)} \right],$$

$$c^2 \mathscr{C}_1 \mathbb{E}\left[\frac{1}{\left(R^2 + \frac{p+k-2}{p-2}\eta^2\right)\left(R^2 + \frac{(p+k-4)(p-14)}{(p-2)(p-4)}\eta^2\right)} \right] - \frac{2c}{p-2} \mathbb{E}\left[\frac{R^4}{\left(R^2 + \frac{p+k}{p-4}\eta^2\right)^2} \right]$$

and the upper and lower bounds of $\Delta\mathscr{R}_R(\theta)$ are respectively

$$c^2 \mathscr{C}_2 \mathscr{C}_3 \mathbb{E}\left[\frac{R^8}{\left[R^2 + \frac{p+K+4}{p-4}\eta^2\right]^2} \right] - 8c\mathscr{C}_3 p \mathbb{E}\left[\frac{R^6}{R^2 + \frac{(p+k+4)(p-4)}{(p-2)^2}\eta^2} \right]$$
$$-2c\mathscr{C}_3 \mathbb{E}\left[\frac{R^8}{\left(R^2 + \frac{p+k+6}{p-2}\eta^2\right)\left(R^2 + \frac{(p+k+4)(p-14)}{(p-2)(p-4)}\eta^2\right)} \right]$$

$$c^2 \mathscr{C}_2 \mathscr{C}_3 \mathbb{E}\left[\frac{R^8}{\left(R^2 + \frac{p+k+6}{p-2}\eta^2\right)\left(R^2 + \frac{(p+k+4)(p-14)}{(p-2)(p-4)}\eta^2\right)} \right]$$
$$-8c\mathscr{C}_3 p \mathbb{E}\left[\frac{R^6}{R^2 + \frac{p+k+2}{p-4}\eta^2} \right] - 2c\mathscr{C}_3 \mathbb{E}\left[\frac{R^8}{\left[R^2 + \frac{p+K+4}{p-4}\eta^2\right]^2} \right]$$

where $\mathscr{C}_1 = \frac{(p+k-2)(p+k-4)}{(p-2)(p-4)}$, $\mathscr{C}_2 = \frac{(k+4)(k+6)}{p-4}$ and $\mathscr{C}_3 = \frac{k(k+2)}{(p-2)(p+k)(p+k+2)}$.

Remark All bounds given above are exact at 0 since they are deduced from bounds of $\mathbb{E}[(U'U)^{\alpha}/X'X]$ and $\mathbb{E}[(U'U)^{\alpha}/(X'X)^2]$ which are also exact at 0.

3.2 Bounds in Terms of Moments of R^2

It is often desirable to have bounds in terms of moments of R^2. An application of the Jensen inequality to the function $\left(R^{2q}/(R^2+A)(R^2+B)\right)$, where q is the fixed integer and A and B are a fixed non-negative constants, leads to the following lemma.

Lemma 3.1 *For any fixed integer $q \in \mathbf{N}^{\star}$ and $(A, B) \in \mathbb{R}_{+}^{2}$, we have*

$$\frac{\mathbb{E}\left[R^{2(q-2)}\right]}{(1+A\mathbb{E}[R^{-2}])\,(1+B\mathbb{E}[R^{-2}])} \leq \mathbb{E}\left[\frac{R^{2q}}{(R^2+A)(R^2+B)}\right]$$
$$\leq \frac{\mathbb{E}[R^{2q}]\left(\mathbb{E}[R^{-2}]\right)^2}{(1+A\mathbb{E}[R^{-2}])\,(1+B\mathbb{E}[R^{-2}])}.$$

and for $q=0$

$$\frac{\left(\mathbb{E}[R^{-2}]\right)^4}{(\mathbb{E}[R^{-2}]+A\mathbb{E}[R^{-4}])\,(\mathbb{E}[R^{-2}]+B\mathbb{E}[R^{-4}])} \leq \mathbb{E}\left[\frac{1}{(R^2+A)(R^2+B)}\right]$$
$$\leq \frac{\mathbb{E}[R^{-4}](\mathbb{E}[R^2])^2}{(\mathbb{E}[R^2]+A)(\mathbb{E}[R^2]+B)}.$$

Knowing the distribution of radius R we can easily calculate the moments of any order when they exist. Hence the interest of the following proposition which provides the bounds of difference in risk $\Delta\mathscr{R}_R(\theta)$ and $\Delta\mathscr{R}(\theta)$ in terms of moments of R. In the sequel we will give bounds for the expectations considered in the Corollaries 3.2 and 3.3 in terms of these moments.

Let us set,

$$A = \frac{(p+k-2)(p+k-4)}{(p-2)(p-4)}, \quad B = \frac{k(k+2)}{(p-2)(p-4)}$$

and

$$C = \frac{k(k+2)(k+4)(k+6)}{(p-2)(p-4)(p+k)(p+k+2)}, \quad D = \frac{k(k+2)(k+4)}{(p-2)(p+k)(p+k+2)}.$$

According to Lemma 3.1, we deduce the following result.

Corollary 3.4 *If $p \geq 14$, then for any $\theta \in \mathbb{R}^p$ we have*

$$\frac{A\left(\mathbb{E}[R^{-2}]\right)^4}{\left(\mathbb{E}[R^{-2}] + \frac{p+k-4}{p-4}\eta^2\right)^2} \leq \mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]$$
$$\leq \frac{A\mathbb{E}[R^{-4}](\mathbb{E}[R^{-2}])^2}{\left(\mathbb{E}[R^2] + \frac{p+k-2}{p-2}\eta^2\right)\left(\mathbb{E}[R^2] + \frac{(p+k-4)(p-14)}{(p-2)(p-4)}\eta^2\right)}$$

$$\frac{B\mathbb{E}[R^2]}{\left(1 + \frac{p+k}{p-4}\eta^2\mathbb{E}[R^{-2}]\right)^2} \leq \mathbb{E}_\theta\left[\frac{||U||^4}{\|X\|^4}\right]$$
$$\leq \frac{B\mathbb{E}[R^4](\mathbb{E}[R^{-2}])^2}{\left(1 + \frac{p+k+2}{p-2}\eta^2\mathbb{E}[R^{-2}]\right)\left(1 + \frac{(p+k)(p-14)}{(p-2)(p-4)}\eta^2\mathbb{E}[R^{-2}]\right)}$$

$$\frac{C\mathbb{E}[R^6]}{\left(1 + \frac{p+k-4}{p-4}\eta^2\mathbb{E}[R^{-2}]\right)^2} \leq \mathbb{E}_\theta\left[\frac{||U||^8}{\|X\|^4}\right]$$
$$\leq \frac{C\mathbb{E}[R^8](\mathbb{E}[R^{-2}])^2}{\left(1 + \frac{p+k+6}{p-2}\eta^2\mathbb{E}[R^{-2}]\right)\left(1 + \frac{(p+k+4)(p-14)}{(p-2)(p-4)}\eta^2\mathbb{E}[R^{-2}]\right)}$$

and

$$\frac{D(\mathbb{E}[R^4])^2}{\mathbb{E}[R^4] + \frac{p+k+2}{p-4}\eta^2\mathbb{E}[R^2]} \leq \mathbb{E}_\theta\left[\frac{||U||^6}{\|X\|^2}\right] \leq \frac{D\mathbb{E}[R^6]\mathbb{E}[R^4]}{\mathbb{E}[R^6] + \frac{(p+k+4)(p-4)}{(p-2)^2}\eta^2\mathbb{E}[R^4]}.$$

The proof of this corollary is obtained by applying the Corollaries 3.2, 3.3, and Lemma 6.2. The last double inequality was obtained in [2]. ■

Proposition 3.5 *If $p \geq 14$, then for any fixed $\theta \in \mathbb{R}^p$, we have*

$$\frac{k(k+2)}{(p-2)(p+k)(p+k+2)}\left\{-8p\frac{\mathbb{E}[R^6]\mathbb{E}[R^4]}{\mathbb{E}[R^6] + \frac{p+k+4}{(p-2)^2}\eta^2\mathbb{E}[R^4]} + \frac{(k+4)(k+6)}{p-4}\right.$$
$$\left.\times\left(c - \frac{2(p-4)}{(k+4)(k+6)}\right)\frac{\mathbb{E}[R^8](\mathbb{E}[R^{-2}])^2}{\left(1 + \frac{p+k+6}{p-2}\eta^2\mathbb{E}[R^{-2}]\right)\left(1 + \frac{(p+k+4)(p-14)}{(p-2)(p-4)}\eta^2\mathbb{E}[R^{-2}]\right)}\right\}$$

$$
\begin{aligned}
&\leq \Delta\mathscr{R}_R(\theta) \\
&\leq \frac{k(k+2)}{(p-2)(p+k)(p+k+2)}\left\{\frac{(k+4)(k+6)}{p-4}\left(c-\frac{2(p-4)}{(k+4)(k+6)}\right)\right. \\
&\times \left.\frac{\mathbb{E}[R^6]}{\left(1+\frac{p+k-4}{p-4}\eta^2\mathbb{E}[R^{-2}]\right)^2}-8p\frac{(\mathbb{E}[R^4])^2}{\mathbb{E}[R^4]+\frac{p+k+2}{p-4}\eta^2\mathbb{E}[R^2]}\right\}.
\end{aligned}
$$

4 Comparison Between Robust and Non-robust Estimators

This section gives sufficient conditions for the optimal robust estimator to dominate any non-robust estimator λ_c.

From (13) it is easy to see that the optimal constant c which minimizes the risk of λ_α^c is

$$
c_\alpha^\star(\eta)=\frac{\frac{p-4}{(k+\alpha)(k+\alpha+2)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right]+\frac{p\alpha}{k+\alpha}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]}{\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]} \tag{15}
$$

and the corresponding estimator $\lambda_\alpha^{c_\alpha^\star}$ has the risk

$$
R(\lambda_u,X,\theta)-\frac{\left(\frac{p-4}{(k+\alpha)(k+\alpha+2)}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right]+\frac{p\alpha}{k+\alpha}\mathbb{E}_\theta\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]\right)^2}{\mathbb{E}_\theta\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]}. \tag{16}
$$

The following theorem gives a sufficient condition for the optimal robust estimator to dominate any non-robust estimator.

Theorem 4.1 *The optimal robust estimator $\lambda_\alpha^{c_\alpha^\star}$ uniformly (in θ) dominates all the non-robust estimator λ_0^c provided that for any $\theta\in\mathbb{R}^p$ and for any $p\geq 5$,*

$$
\frac{\mathbb{E}\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]\left(\mathbb{E}\left[\frac{\|U\|^4}{\|X\|^4}\right]\right)^2}{\mathbb{E}\left[\frac{1}{\|X\|^4}\right]\left(\mathbb{E}\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right]+p\alpha\frac{k+\alpha+2}{p-4}\mathbb{E}\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]\right)^2}\leq\left(\frac{k(k+2)}{(k+\alpha)(k+\alpha+2)}\right)^2. \tag{17}
$$

Proof According to (15) the optimal choice of the constant c leading to a minimum risk for non-robust estimator λ_0^c depend on θ and equals to

$$
c_0^\star=\frac{p-4}{k(k+2)}\frac{\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^2}\right]}{\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]}, \tag{18}
$$

the corresponding estimator $\lambda_0^{c_0^\star}$ has the risk

$$R(\lambda_0^{c_0^\star}, X, \theta) = R(\lambda_u, X, \theta) - \left(\frac{p-4}{k(k+2)}\right)^2 \frac{\left(\mathbb{E}_\theta\left[\frac{\|U\|^4}{\|X\|^4}\right]\right)^2}{\mathbb{E}_\theta\left[\frac{1}{\|X\|^4}\right]}. \tag{19}$$

According to (18) and (19), the risk difference $\Delta R(\theta)$ between $\lambda_0^{c_0^\star}$ and $\lambda_4^{c_4^\star}$ is

$$\begin{aligned}\Delta R(\theta) &= R(\lambda_4^{c_4^\star}, X, \theta) - R(\lambda_0^{c_0^\star}, X, \theta)\\ &= \frac{\left(\frac{p-4}{(k)(k+2)}\right)^2 \left(\mathbb{E}\left[\frac{\|U\|^4}{\|X\|^4}\right]\right)^2}{\mathbb{E}\left[\frac{1}{\|X\|^4}\right]}\\ &\quad - \frac{\left(\frac{p-4}{(k+\alpha)(k+\alpha+2)}\right)^2 \left(\mathbb{E}\left[\frac{\|U\|^{\alpha+4}}{\|X\|^4}\right] + \frac{p\alpha(k+\alpha+2)}{p-4}\mathbb{E}\left[\frac{\|U\|^{\alpha+2}}{\|X\|^2}\right]\right)^2}{\mathbb{E}\left[\frac{\|U\|^{2\alpha}}{\|X\|^4}\right]}.\end{aligned}$$

Thus $\lambda_4^{c_4^\star}$ uniformly dominates all the estimator λ_0^c (for the same c), if this last quantity is negative, that is, if which is the desired result. ■

The condition (17), may be difficult to verify directly and a convenient way is to express the upper bound of the left-hand side (17) using the bound obtained in Sect. 3. This is the main idea of the following corollary.

Corollary 4.1 *For $p \geq 14$, a sufficient condition for which $\lambda_\alpha^{c_\alpha^\star}$ dominates uniformly (in θ) all the estimator λ_0^c is*

$$\frac{\mathbb{E}\left[\frac{R^{2\alpha}}{\left(R^2+\frac{p+k+2\alpha-2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k+2\alpha-4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]\left(\mathbb{E}\left[\frac{R^4}{\left(R^2+\frac{p+k+2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]\right)^2}{\mathbb{E}\left[\frac{1}{\left(R^2+\frac{p+k-4}{p-4}\|\theta\|^2\right)^2}\right]\left(\mathbb{E}\left[\frac{R^{\alpha+4}}{\left(R^2+\frac{p+k+\alpha}{p-4}\|\theta\|^2\right)^2}\right] + p\alpha\mathbb{E}\left[\frac{R^{\alpha+2}}{R^2+\frac{p+k+\alpha-2}{p-4}\|\theta\|^2}\right]\right)^2}$$

$$\leq \frac{\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{p+k}{2}+\alpha\right)\Gamma^2\left(\frac{p+k+2}{2}\right)(p+k-2)(p+k-4)(p+k+\alpha)^2(k+\alpha)^2}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{k}{2}+\alpha\right)\Gamma^2\left(\frac{p+k+\alpha+2}{2}\right)(p+k+2\alpha-2)(p+k+2\alpha-4)}$$

for $\alpha = 4$

$$\frac{\mathbb{E}\left[\frac{R^8}{\left(R^2+\frac{p+k+6}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k+4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]\left(\mathbb{E}\left[\frac{R^4}{\left(R^2+\frac{p+k+2}{p-2}\|\theta\|^2\right)\left(R^2+\frac{(p+k)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}\right]\right)^2}{\mathbb{E}\left[\frac{1}{\left(R^2+\frac{p+k-4}{p-4}\|\theta\|^2\right)^2}\right]\left(\mathbb{E}\left[\frac{R^8}{\left(R^2+\frac{p+k+4}{p-4}\|\theta\|^2\right)^2}\right]+4p\mathbb{E}\left[\frac{R^6}{R^2+\frac{p+k+2}{p-4}\|\theta\|^2}\right]\right)^2}$$
$$\leq \frac{k(k+2)(p+k-2)(p+k-4)}{(k+4)(k+6)(p+k)(p+k+2)}. \tag{20}$$

It is interesting to first consider the condition (20) for $\theta = 0$ and for η at infinity.

Corollary 4.2 *The optimal robust estimator* $\lambda_4^{c_4^\star}$ *dominates all the estimators* λ_0^c *at* $\theta = 0$ *if:*

$$\mathbb{E}\left[\frac{1}{R^4}\right]\mathbb{E}\left[R^4\right] \geq \frac{(k+4)(k+6)(p+k)(p+k+2)}{k(k+2)(4p+1)^2(p+k-2)(p+k-4)}. \tag{21}$$

Proof The result is a straightforward application of Corollary 4.1 for $\theta = 0$. ■

The domination condition of $\lambda_4^{c_4^\star}$ over λ_0^c, for $\eta = \|\theta\|$ at infinity, is deduced by dividing the numerator and denominator of the left-hand side of (20) by η^4.

Corollary 4.3 *The optimal robust estimator* $\lambda_4^{c_4^\star}$ *dominates all the estimators* λ_0^c *for* η *at infinity if*

$$\frac{\mathbb{E}\left[R^8\right]\mathbb{E}\left[R^4\right]^2}{\left(\frac{p-4}{(p+k+4)^2}\mathbb{E}\left[R^8\right]+\frac{4p}{p+k+2}\mathbb{E}\left[R^6\right]\right)^2}$$
$$\leq \frac{k(k+2)(p+k)(p+k+2)(p+k+4)(p+k+6)(p+k-2)(p-14)}{(p-2)^3(p+k-4)(k+4)(k+6)}. \tag{22}$$

5 Examples

As in the case of the point estimate, the class of Student distribution provides a framework where the condition of domination of Theorem 4.1 is achieved. Suppose that $(X, U)'$ has a Student distribution with m degrees of freedom. Straightforward calculations show that the density of the radial distribution is given by

$$\frac{2\Gamma\left(\frac{m+p+k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p+k}{2}\right)m^{\frac{p+k}{2}}}\left(1+\frac{R^2}{m}\right)^{-\frac{p+k+m}{2}} R^{p+k-1}$$

The key terms in the various proofs of the results are

$$s(\gamma, A, B) = \mathbb{E}\left[\frac{R^{\gamma}}{(R^2 + A\eta^2)\,(R^2 + B\eta^2)}\right]$$

where A, B, and γ to be positive constant. The following lemma gives an upper bound of this term and a lower bound in the case where $A = B$.

Lemma 5.1 *For $A > 0$, $B > 0$ and $\gamma > 0$ fixed, we have*

$$s(\gamma, A, B) \leq \frac{m^{\frac{p+k+\gamma-4}{2}}}{2}\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma-4}{2}\right)$$

and

$$\frac{m^{\frac{p+k+\gamma-4}{2}}(p+k+m+4)^2\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}\right)}{2A^2(m+4-\gamma)^2\left(\frac{p+k+\gamma}{A(m+4-\gamma)} + \frac{\eta^2}{m}\right)^2} \leq s(\gamma, A, A)$$

Whenever the dimension k of the residual vector U is large enough, the following proposition gives a sufficient condition to the dominance.

Proposition 5.1 *For $p \geq 14$ and for $m \geq 5$, there exist $k_0 > 0$ such that for any $k \geq k_0$ the optimal robust estimator $\lambda_R^{c_4^*}(X, U)$ dominates uniformly in θ all non-robust estimators $\lambda_c(X)$.*

Proof Notice that the left side of inequality (17) is dominated by $I \times J$ where

$$I = \frac{\mathbb{E}\left[\frac{R^8}{\left(R^2+\mathscr{A}_1\eta^2\right)\left(R^2+\mathscr{A}_2\eta^2\right)}\right]}{\mathbb{E}\left[\frac{R^8}{\left(R^2+\mathscr{A}_5\eta^2\right)^2}\right]}$$

and

$$J = \frac{\left(\mathbb{E}\left[\frac{R^4}{\left(R^2+\mathscr{A}_3\eta^2\right)\left(R^2+\mathscr{A}_4\eta^2\right)}\right]\right)^2}{\mathbb{E}\left[\frac{1}{\left(R^2+\mathscr{A}_6\eta^2\right)^2}\right]\mathbb{E}\left[\frac{R^8}{\left(R^2+\mathscr{A}_5\eta^2\right)^2}\right]}.$$

Firstly, we have

$$I \leq \left(\frac{m-4}{p+k+m+4}\right)^2\left(\frac{p+k+8}{m-4} + \mathscr{A}_5\frac{\eta^2}{m}\right)^2\frac{(p+k+m)(p+k+m+2)}{(p+k+4)(p+k+6)}. \tag{23}$$

Indeed

$$I \le \frac{m^{\frac{p+k+4}{2}}\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2}\right)\int_0^1 \frac{1}{t^2}\frac{\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2},dt\right)}{\left(\frac{1-t}{t}+\mathscr{A}_1\frac{\eta^2}{m}\right)\left(\frac{1-t}{t}+\mathscr{A}_2\frac{\eta^2}{m}\right)}}{m^{\frac{p+k+4}{2}}\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2}\right)\int_0^1 \frac{1}{t^2}\frac{\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2},dt\right)}{\left(\frac{1-t}{t}+\mathscr{A}_5\frac{\eta^2}{m}\right)^2}}$$

$$\le \int_0^1 \frac{1}{t^2}\frac{\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2},dt\right)}{\left(\frac{1-t}{t}+\mathscr{A}_1\frac{\eta^2}{m}\right)\left(\frac{1-t}{t}+\mathscr{A}_2\frac{\eta^2}{m}\right)}\left(\frac{1}{1-\frac{\mathscr{B}\left(\frac{m-4}{2}+1,\frac{p+k+8}{2}\right)}{\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2}\right)}\left(1-\mathscr{A}_5\frac{\eta^2}{m}\right)}\right)^{-2}$$

$$= \left(\frac{m-4}{p+k+m+4}\right)^2\left(\frac{p+k+8}{m-4}+\mathscr{A}_5\frac{\eta^2}{m}\right)^2\frac{(p+k+m)(p+k+m+2)}{(p+k+4)(p+k+6)}.$$

On the other hand, we have

$$J \le \frac{\mathscr{A}_5^2\mathscr{A}_6^2(m+4)^2(m-4)^2(p+k+m)(p+k+m+2)(m-4)(p+k+m)}{m(p+k+m+4)^4\mathscr{A}_3^2\mathscr{A}_4^2(m+2)(p+k)(p+k+2)(p+k+4)} \times\frac{(m-2)(p+k+m+2)}{(p+k+6)}\left[\frac{p+k}{\mathscr{A}_5(m+4)}+\frac{\eta^2}{m}\right]^2\left[\frac{p+k+8}{\mathscr{A}_6(m-4)}+\frac{\eta^2}{m}\right]^2 \quad (24)$$

where

$$\mathscr{A}_1=\frac{p+k+6}{p-2},\quad \mathscr{A}_2=\frac{(p+k+4)(p-14)}{(p-2)(p-4)},\quad \mathscr{A}_3=\frac{p+k+2}{p-2}$$

$$\mathscr{A}_4=\frac{(p+k)(p-14)}{(p-2)(p-4)},\quad \mathscr{A}_5=\frac{p+k-4}{p-4}\ \text{and}\ \mathscr{A}_6=\frac{p+k+4}{p-4}.$$

Indeed

$$J \le \frac{m^{p+k}\mathscr{B}^2\left(\frac{m}{2},\frac{p+k+4}{2}\right)\left(\int_0^1\frac{1}{t^2}\frac{\mathscr{B}\left(\frac{m}{2},\frac{p+k+4}{2},dt\right)}{\left(\frac{1-t}{t}+\mathscr{A}_3\frac{\eta^2}{m}\right)\left(\frac{1-t}{t}+\mathscr{A}_4\frac{\eta^2}{m}\right)}\right)^2}{\frac{m+p+k(p+k+m+4)^4\mathscr{B}\left(\frac{m+4}{2},\frac{p+k}{2}\right)\mathscr{B}\left(\frac{m-4}{2},\frac{p+k+8}{2}\right)}{\mathscr{A}_5^2\mathscr{A}_6^2(m+4)^2(m-4)^2\left[\frac{p+k}{\mathscr{A}_5(m+4)}+\frac{\eta^2}{m}\right]\left[\frac{p+k+8}{\mathscr{A}_6(m-4)}+\frac{\eta^2}{m}\right]}}$$

$$\le \frac{\mathscr{A}_5^2\mathscr{A}_6^2(m+4)^2(m-4)^2(p+k+m)(p+k+m+2)(m-4)(m-2)}{\mathscr{A}_3^2\mathscr{A}_4^2 m(p+k+m+4)^4(m+2)(p+k)(p+k+2)(p+k+4)} \times\frac{(p+k+m)(p+k+m+2)}{(p+k+6)}\left[\frac{p+k}{\mathscr{A}_5(m+4)}+\frac{\eta^2}{m}\right]^2\left[\frac{p+k+8}{\mathscr{A}_6(m-4)}+\frac{\eta^2}{m}\right]^2.$$

Thus $I \times J$ is bounded by

$$\frac{\mathscr{A}_5^2\mathscr{A}_6^2(m+4)^2(m-4)^4(p+k+m)(p+k+m+2)(m-4)(m-2)(p+k+m)^2}{\mathscr{A}_3^2\mathscr{A}_4^2 m(p+k+m+4)^6(m+2)(p+k)(p+k+2)(p+k+4)^2}$$
$$\frac{(p+k+m+2)^2}{(p+k+6)^2}\left[\frac{p+k}{\mathscr{A}_5(m+4)}+\frac{\eta^2}{m}\right]^2\left[\frac{p+k+8}{\mathscr{A}_6(m-4)}+\frac{\eta^2}{m}\right]^2$$
$$\left(\frac{p+k+8}{m-4}+\mathscr{A}_5\frac{\eta^2}{m}\right)^2. \tag{25}$$

Remark that when k tends to infinity, the latter term converges to 0 and the right-hand side of Eq. (25) converges to 1. The proof is complete. ∎

Acknowledgements This work was supported by the Hassan II Academy of Sciences and Technology, Rabat, Morocco.

Appendix

The following lemma can be found in [2] to whom we refer for a proof.

Lemma 6.1 *Assume that $p \geq 0$. If g and h are two measurable real-valued functions, then for any $R > 0$ and any $\theta \in \mathbb{R}^p$*

$$\mathbb{E}_{R,\theta}\left[g(X'X).h(U'U)\right] = \int_0^1 h\left(R^2(1-u)\right)\int_0^1 \frac{1}{2}\left[g\left(R^2u+\|\theta\|^2-2\|\theta\|Ru^{\frac{1}{2}}v^{\frac{1}{2}}\right)\right.$$
$$\left.+\,g\left(R^2u+\|\theta\|^2+2\|\theta\|Ru^{\frac{1}{2}}v^{\frac{1}{2}}\right)\right]\times B\left(\frac{1}{2},\frac{p-1}{2},dv\right)B\left(\frac{p}{2},\frac{k}{2},du\right)$$

provided these expectations exist.

Lemma 6.2 *Let F be a probability distribution on $[0,1]$ and a, b two positive constants. Then for any $x \geq 0$, we have*

$$\int_0^1 \frac{1}{t^s}\frac{a+x}{b\frac{1-t}{t}+x}dF(t) \leq \int_{t_1}^0 \frac{1}{t^s}dF(t)+\frac{a}{b}\int_1^{t_1}\frac{t^{1-s}}{1-t}dF(t) \leq \mathbb{E}\left[\frac{1}{t^s}\right]+\frac{a}{b}\mathbb{E}\left[\frac{t^{1-s}}{1-t}\right]$$

where t_1 is such that $\frac{a}{b}\frac{t_1}{1-t_1}=1$ (i.e., $t_1=\left(1+\frac{b}{a}\right)^{-1}$) and E is the expectation with respect to F.

Proof Notice that for $0 \leq t \leq t_1$ (respectively $t_1 \leq t \leq 1$) the mapping $x \mapsto (a+x)/(b\frac{1-t}{t}+x)$ is increasing (respectively decreasing). The conclusion of the lemma follows by bounding this function by its value at infinity (respectively at 0). ∎

Proof of the Proposition 3.1 According to Lemma 1 given in [2] applied to $g(x'x) = (x'x)^{-1}$ and $h(u'u) = (u'u)^q$ gives, for $R \geq 0$ and $\theta \in \mathbf{R}^p$,

$$\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] = R^{2q}\int_0^1\int_0^1 \frac{(R^2u+\|\theta\|^2)^2+4R^2\|\theta\|^2uv}{((R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv)^2}(1-u)^q \times B\left(\frac{1}{2},\frac{p-1}{2},dv\right)B\left(\frac{p}{2},\frac{k}{2},du\right).$$

Inserting $(1-u)^q$ into the beta distribution $B(p/2,k/2,du)$, the last expectation becomes

$$\frac{\Gamma\left(\frac{k+p}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{k+p}{2}+q\right)\Gamma\left(\frac{k}{2}\right)}R^{2q} \times\int_0^1\int_0^1 \frac{(R^2u+\|\theta\|^2)^2+4R^2\|\theta\|^2uv}{((R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv)^2}B\left(\frac{1}{2},\frac{p-1}{2},dv\right)B\left(\frac{p}{2},\frac{k}{2}+q,du\right).$$

■

Proof of the Proposition 3.2 Let $R \geq 0$ and θ be fixed in $\mathbf{R}^p$. From the expression of $\mathbb{E}_\theta\left[\|U\|^q/\|X\|^2\right]$ given by Proposition 2.1, it is clear that the integral terms between brackets is equal to

$$\int_0^1 \frac{1}{(R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv}+\frac{8R^2\|\theta\|^2uv}{((R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv)^2}B\left(\frac{1}{2},\frac{p-1}{2},dv\right)$$

$$= \frac{p-2}{p-3}\int_0^1 \frac{1-v}{(R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2v}B\left(\frac{1}{2},\frac{p-3}{2},dv\right) +8R^2\|\theta\|^2u\frac{p-2}{p-3}\int_0^1\frac{v(1-v)}{\left((R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2v\right)^2}B\left(\frac{1}{2},\frac{p-1}{2},dv\right).$$

Since the functions $v \to (1-v)\left[\left(R^2u+\|\theta\|^2\right)^2-4R^2u\|\theta\|^2v\right]^{-1}$ is concave and non-increasing and the function $v \to v\left[\left(R^2u+\|\theta\|^2\right)^2-4R^2u\|\theta\|^2v\right]^{-1}$ is non-decreasing then by Jensen's and covariance inequalities, the first and the second terms in the least equality are bounded by

$$\frac{p-2}{p-3}\frac{\frac{p-2}{p-3}}{(R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2+4R^2u\|\theta\|^2\frac{p-3}{p-2}} +8R^2\|\theta\|^2u\frac{p-2}{p-3}\int_0^1\frac{1-v}{(R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2v}B\left(\frac{1}{2},\frac{p-3}{2},dv\right)$$

$$
\begin{aligned}
&\times \int_0^1 \frac{v}{(R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2 v} B\left(\frac{1}{2},\frac{p-3}{2},dv\right).\\
&\le \frac{1}{(R^2u+\|\theta\|^2)^2-\frac{4}{p-2}R^2u\|\theta\|^2}+\frac{8R^2\|\theta\|^2u}{(R^2u+\|\theta\|^2)^2-\frac{4}{p-2}R^2u\|\theta\|^2}\\
&\times\frac{1}{p-2}\frac{p-2}{p-5}\frac{\frac{p-5}{p-2}}{(R^2u+\|\theta\|^2)^2-4R^2u\|\theta\|^2+4R^2u\|\theta\|^2\frac{p-5}{p-2}}\\
&= \frac{1}{(R^2u+\|\theta\|^2)^2-\frac{12}{p-2}R^2u\|\theta\|^2}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\int_0^1\int_0^1 \frac{(R^2u+\|\theta\|^2)^2+4R^2\|\theta\|^2uv}{((R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv)^2} B\left(\frac{1}{2},\frac{p-1}{2},dv\right)B\left(\frac{p}{2},\frac{k}{2}+q,du\right)\\
&\le \frac{B\left(\frac{p}{2}-1,\frac{k}{2}+q\right)}{B\left(\frac{p}{2},\frac{k}{2}+q\right)}\int_0^1 \frac{u}{R^2u+\|\theta\|^2}\frac{1}{R^2u+\frac{p-14}{p-2}\|\theta\|^2}B\left(\frac{p}{2}-1,\frac{k}{2}+q,du\right).
\end{aligned}
$$

Since the function $u \mapsto u\left(R^2u+\|\theta\|^2\right)^{-1}$ and $u \mapsto \left(R^2u+\frac{p-14}{p-2}\|\theta\|^2\right)^{-1}$ are respectively non-decreasing and non-increasing then by covariance inequality the least equality is bounded by

$$
\begin{aligned}
&\frac{B\left(\frac{p}{2}-1,\frac{k}{2}+q\right)}{B\left(\frac{p}{2},\frac{k}{2}+q\right)}\int_0^1 \frac{u}{R^2u+\|\theta\|^2}B\left(\frac{p}{2}-1,\frac{k}{2}+q,du\right)\\
&\times\int_0^1 \frac{1}{R^2u+\frac{p-14}{p-2}\|\theta\|^2}B\left(\frac{p}{2}-1,\frac{k}{2}+q,du\right)\\
&= \frac{B\left(\frac{p}{2}-1,\frac{k}{2}+q\right)}{B\left(\frac{p}{2},\frac{k}{2}+q\right)}\int_0^1 \frac{u}{R^2u+\|\theta\|^2}B\left(\frac{p}{2}-1,\frac{k}{2}+q,du\right)\\
&\times\frac{B\left(\frac{p}{2}-2,\frac{k}{2}+q\right)}{B\left(\frac{p}{2}-1,\frac{k}{2}+q\right)}\int_0^1 \frac{u}{R^2u+\frac{p-14}{p-2}\|\theta\|^2}B\left(\frac{p}{2}-1,\frac{k}{2}+q,du\right).
\end{aligned}
$$

The mappings $u \mapsto u\left(R^2u+\|\theta\|^2\right)^{-1}$ and $u \mapsto u\left(R^2u+\frac{p-14}{p-2}\|\theta\|^2\right)^{-1}$ are concave, then by Jensen's inequality the least equality is bounded by

$$
\frac{1}{R^2\frac{p-2}{p+k+2q-2}+\|\theta\|^2}\times\frac{1}{R^2\frac{p-4}{p+k+2q-4}+\frac{p-14}{p-2}\|\theta\|^2}
$$

$$= \frac{\frac{(p+k+2q-2)(p+k+2q-4)}{(p-2)(p-4)}}{\left(R^2 + \frac{p+k+2q-2}{p-2}\|\theta\|^2\right)\left(R^2 + \frac{(p+k+2q-4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}$$

Henceforth

$$\mathbb{E}_{R,\theta}\left[\frac{||u||^{2q}}{||X||^4}\right] \leq \frac{\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{p+k}{2}+q\right)\Gamma\left(\frac{k}{2}\right)} \frac{(p+k+2q-2)(p+k+2q-4)}{(p-2)(p-4)}$$
$$\frac{R^{2q}}{\left(R^2 + \frac{p+k+2q-2}{p-2}\|\theta\|^2\right)\left(R^2 + \frac{(p+k+2q-4)(p-14)}{(p-2)(p-4)}\|\theta\|^2\right)}$$

■

Proof of the Proposition 3.3 Let $R \geq 0$ and θ be fixed in $\mathbf{R}^p$, The Proposition 2.1 yields

$$\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] = \frac{B\left(\frac{p}{2}, \frac{k}{2}+q\right)}{B\left(\frac{p}{2}, \frac{k}{2}\right)} R^{2q}\left(k_1 + k_2\right) \tag{26}$$

where

$$k_1 = \int_0^1 \int_0^1 \frac{B\left(\frac{1}{2}, \frac{p-1}{2}, dv\right) B\left(\frac{p}{2}, \frac{k}{2}+q, du\right)}{(R^2u + \|\theta\|^2)^2 - 4R^2\|\theta\|^2 uv}.$$

It is clear that

$$\int_0^1 \frac{B\left(\frac{1}{2}, \frac{p-1}{2}, dv\right)}{(R^2u + \|\theta\|^2)^2 - 4R^2\|\theta\|^2 uv} \geq \frac{1}{(R^2u + \|\theta\|^2)^2}$$

hence

$$k_1 \geq \int_0^1 \frac{1}{R^2u + \|\theta\|^2} \frac{1}{R^2u + \|\theta\|^2} B\left(\frac{p}{2}, \frac{k}{2}+q, du\right).$$

By covariance inequality we have

$$k_1 \geq \left[\int_0^1 \frac{B\left(\frac{p}{2}, \frac{k}{2}+q, du\right)}{R^2u + \|\theta\|^2}\right]^2.$$

Furthermore, by Jensen's inequality, we obtain

$$\int_0^1 \frac{B\left(\frac{p}{2},\frac{k}{2}+q,du\right)}{R^2u+\|\theta\|^2} \geq \frac{\frac{p+k+2q-4}{p-4}}{R^2+\frac{p+k+2q-4}{p-4}||\theta||^2}$$

since the function $u \to \left(R^2u+||\theta||^2\right)^{-1}$ is convex. Finally we have obtained the lower bound of k_1.

By same argument we get a lower bound of k_2

$$\begin{aligned}
k_2 &= \int_0^1\int_0^1 \frac{8R^2\|\theta\|^2uv}{((R^2u+\|\theta\|^2)^2-4R^2\|\theta\|^2uv)^2} B\left(\frac{1}{2},\frac{p-1}{2},dv\right) B\left(\frac{p}{2},\frac{k}{2}+q,du\right) \\
&\geq \int_0^1 \frac{8R^2\|\theta\|^2u}{(R^2u+\|\theta\|^2)^4} B\left(\frac{p}{2},\frac{k}{2}+q,du\right)\int_0^1 vB\left(\frac{1}{2},\frac{p-1}{2},dv\right) \\
&\geq \frac{8R^2\|\theta\|^2}{p}\frac{\frac{p}{2}}{\frac{p+k+2q}{2}}\left[\frac{B\left(\frac{p-2}{2},\frac{k}{2}+q\right)}{B\left(\frac{p+2}{2},\frac{k}{2}+q\right)}\int_0^1 \frac{u^2}{R^2u+\|\theta\|^2}B\left(\frac{p-2}{2},\frac{k}{2}+q,du\right)\right]^4 \\
&\geq \frac{8R^2\|\theta\|^2}{p+k+2q}\left[\frac{1}{\frac{p-2}{p+k+2q-2}R^2+\|\theta\|^2}\right]^4.
\end{aligned}$$

Consequently, we obtain a lower bound of $\mathbb{E}_{R,\theta}\left[\|U\|^{2q}/\|X\|^4\right]$

$$\mathbb{E}_{R,\theta}\left[\frac{\|U\|^{2q}}{\|X\|^4}\right] \geq \frac{\Gamma\left(\frac{p+k}{2}\right)\Gamma\left(\frac{k}{2}+q\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{p+k}{2}+q\right)}\left[\frac{R^{2q}}{\left[\frac{p-4}{p+K+2q-4}R^2+\|\theta\|^2\right]^2}\right].$$

We obtain the desired result by unconditional. ■

Proof of the Lemma 3.1 Let $A \geq 0$ and $B \geq 0$ fixed. For $q \geq 1$, the function $R^{2q}/(R^2+A)$ is non-decreasing in R and $1/(R^2+B)$ is decreasing in R. The covariance inequality gives

$$\mathbb{E}\left[\frac{R^{2q}}{(R^2+A)(R^2+B)}\right] \leq \mathbb{E}\left[\frac{R^{2q}}{R^2+A}\right]\mathbb{E}\left[\frac{1}{R^2+B}\right] \leq \frac{\mathbb{E}[R^{2q}]\left(\mathbb{E}[R^{-2}]\right)^2}{(1+A\mathbb{E}[R^{-2}])\,(1+B\mathbb{E}[R^{-2}])}.$$

Furthermore it is easy to show by same argument that

$$\mathbb{E}\left[\frac{R^{2q}}{(R^2+A)(R^2+B)}\right]\frac{\mathbb{E}\left[R^{2(q-2)}\right]}{(1+A\mathbb{E}[R^{-2}])\,(1+A\mathbb{E}[R^{-2}])}$$

For $q = 0$, we have

$$\mathbb{E}\left[\frac{1}{(R^2+A)(R^2+B)}\right]\mathbb{E}[R^{-4}]\mathbb{E}\left[\frac{R^2}{R^2+A}\right]\mathbb{E}\left[\frac{R^2}{R^2+B}\right] \geq \frac{\mathbb{E}[R^{-4}](\mathbb{E}[R^2])^2}{(\mathbb{E}[R^2]+A)(\mathbb{E}[R^2]+B)}$$

and that

$$\mathbb{E}\left[\frac{1}{(R^2+A)(R^2+B)}\right] \geq \frac{\left(\mathbb{E}[R^{-2}]\right)^4}{\left(\mathbb{E}[R^{-2}]+A\mathbb{E}[R^{-4}]\right)\left(\mathbb{E}[R^{-2}]+B\mathbb{E}[R^{-4}]\right)}.$$

■

Prof of the Lemma 5.1 Let A and B two fixed positive constants, then

$$s(\gamma, A, B) = \int_0^\infty \frac{\left(1+\frac{R^2}{m}\right)^{-\frac{p+k+m}{2}} R^{p+k+\gamma-1}}{(R^2+A\eta^2)(R^2+B\eta^2)} dR$$

making a change of variable $t = \left(1+\frac{R^2}{m}\right)^{-1}$, we get $R^2 = m(\frac{1}{t}-1)$ and $dR = -\frac{m^{1/2}}{2}t^{-3/2}(1-t)^{-1/2}$. Therefore $s(\gamma, A, B)$ is equal to

$$\frac{m^{\frac{p+k+\gamma-4}{2}}}{2}\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}\right)\int_0^1 \frac{\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}, dt\right)}{\left[1-t\left(1-\frac{A\eta^2}{m}\right)\right]\left[1-t\left(1-\frac{B\eta^2}{m}\right)\right]}. \tag{27}$$

For $A = B$, we have by Jensen's inequality

$$\int_0^1 \frac{\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}, dt\right)}{\left[1-t\left(1-\frac{A\eta^2}{m}\right)\right]^2} \geq \left[1-\frac{\mathscr{B}\left(\frac{m+4-\gamma}{2}+1, \frac{p+k+\gamma}{2}\right)}{\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}\right)}\left(1-\frac{A\eta^2}{m}\right)\right]^{-2}$$

the last inequalities were derived from the convexity of the functions $t \to t^2$ and $t \to (1-zt)^{-1}$. Noticing that

$$\frac{\mathscr{B}\left(\frac{m+4-\gamma}{2}+1, \frac{p+k+\gamma}{2}\right)}{\mathscr{B}\left(\frac{m+4-\gamma}{2}, \frac{p+k+\gamma}{2}\right)} = \frac{m+4-\gamma}{p+k+m+4}$$

we obtain a lower bound of the last term of the equality (27), let

$$\frac{m^{\frac{p+k+\gamma-4}{2}}(p+k+m+4)^2\mathscr{B}\left(\frac{m+4-\gamma}{2},\frac{p+k+\gamma}{2}\right)}{2E^2(m+4-\gamma)^2\left(\frac{p+k+\gamma}{E(m+4-\gamma)}+\frac{\eta^2}{m}\right)^2}$$

which gives the first inequality of the lemma.

For the second inequality, notice that, from Eq. (27), $s(\gamma, A, B)$ can also be expressed as

$$\begin{aligned} s(\gamma,A,B) &= \frac{m^{\frac{p+k+\gamma-4}{2}}}{2}\mathscr{B}\left(\frac{m+4-\gamma}{2},\frac{p+k+\gamma-4}{2}\right) \\ &\times\int_0^1 \frac{\frac{1-t}{t}}{\left(\frac{1-t}{t}+A\frac{\eta^2}{m}\right)}\frac{\frac{1-t}{t}}{\left(\frac{1-t}{t}+B\frac{\eta^2}{m}\right)}\mathscr{B}\left(\frac{m+4-\gamma}{2},\frac{p+k+\gamma-4}{2},dt\right) \\ &\leq \frac{m^{\frac{p+k+\gamma-4}{2}}}{2}\mathscr{B}\left(\frac{m+4-\gamma}{2},\frac{p+k+\gamma-4}{2}\right). \end{aligned}$$

since, for all $x \in \mathbf{R}^+$, the function $t \to \left(\frac{1-t}{t}\right) / \left(\frac{1-t}{t}+x\right)$ is positive and decreasing, then we get this bound for t tends to zero. ■

References

1. Casella, G., Hwang, J.T.: Limit expressions for the risk of James Stein estimators. Can. J. Stat. **10**, 305–309 (1982)
2. Fourdrinier, D., Strawderman, W.E.: A paradox concerning shrinkage estimators: should a known scale parameter be replaced by an estimated value in the shrinkage factor? J. Multivar. Anal. **59**(2), 109–140 (1996)
3. Fourdrinier, D., Wells, M.T.: Estimation of a loss function for spherically symmetric distribution in the general linear model. Ann. Stat. **23**(2), 571–592 (1995)
4. Fourdrinier, D., Ouassou, I., Strawderman, W.: Estimation of a components vector when some parameters are restricted. J. Multivar. Anal. **86**, 14–24 (2003)
5. Johnstone, I.: On inadmissibility of some unbiased estimates of loss. In: Gupta, S.S., Berger, J.O. (eds.) Statistical Decision Theory and Related Topics IV, vol. 1, pp. 361–379. Academic Press, New York (1988)
6. Lehmann, E.L., Scheffé, H.: Completeness, similar regions and unbiased estimates. Sankhyā, **10**, 305–340 (1950)
7. Lele, C.: Admissibility results in loss estimation. Ann. Stat. **21**, 378–390 (1993)
8. Ouassou, I., Rachdi, M.: Stein type estimation of the regression operator for functional data. Adv. Appl. Stat. Sci. **1**(2), 233–250 (2009)
9. Ouassou, I., Rachdi, M.: Quadratic loss estimation of a location parameter when a subset of its components is unknown. Afr. Stat. **8**(01), 561–575 (2013)

Part II
Applications (Signal Analysis, Epidemiology, Queuing Theory)

Impact of Nonparametric Density Estimation on the Approximation of the $G/G/1$ Queue by the $M/G/1$ One

Aïcha Bareche and Djamil Aïssani

Abstract In this paper, we show the interest of nonparametric boundary density estimation to evaluate a numerical approximation of $G/G/1$ and $M/G/1$ queueing systems using the strong stability approach when the general arrivals law G in the $G/G/1$ system is unknown. A numerical example is provided to support the results. We give a proximity error between the arrival distributions and an approximation error on the stationary distributions of the quoted systems.

1 Introduction

Because of the complexity of some queueing models, analytic results are generally difficult to obtain or are not very exploitable in practice. That is the case, for example, in the $G/G/1$ queueing system, where the Laplace transform or the generating function of the waiting time distribution is not available in a closed form [20]. Indeed, when a practical study is performed in queueing theory, one often replaces a real system by another one which is close to it in some sense but simpler in structure and/or components. The queueing model so constructed represents an idealization of the real queueing one, and hence the "stability" problem arises.

One of the stability methods is the strong stability approach [2, 19] which has been developed in the beginning of the 1980s. It can be used to investigate the ergodicity and stability of the stationary and non-stationary characteristics of Markov chains. In contrast to other methods, the strong stability approach supposes that the perturbation of the transition kernel is small with respect to a certain norm. Such a stringent condition allows us to obtain better estimates on the characteristics of the perturbed chain. Besides the ability to make qualitative analysis of some

A. Bareche (✉)
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Technology, University of Bejaia, Bejaia 06000, Algeria
e-mail: aicha_bareche@yahoo.fr

D. Aïssani
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Exact Sciences, University of Bejaia, Bejaia 06000, Algeria
e-mail: lamos_bejaia@hotmail.com

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_7

complex systems, one importance particularity of the strong stability method is the possibility to obtain stability inequalities with an exact computation of the constants.

The applicability of this approach is well proved and documented in various situations and for different proposals. In particular, it has been applied to several queueing models [1, 10, 15, 17, 23] and inventory models [22].

Note that the first attempt to "measure" the performance of the strong stability method has been used in practice, and has been particularly applied to a simple system of queues [12, 13]. The approach proposed is based on the classical approximation method where the authors perform the numerical proximity of the stationary distribution of an $Hyp/M/1$ (respectively $M/Cox2/1$) system by the one of an $M/M/1$ system when applying the strong stability method. For the first time, Bareche and Aïssani [5] specify an approximation error on the stationary distributions of the $G/M/1$ (resp. $M/G/1$) and $M/M/1$ systems when the general law of arrivals (resp. service times) G is unknown and its density function is estimated by using the kernel density method. In [11], the authors use the discrete event simulation approach and the Student test to measure the performance of the strong stability method through simple numerical examples for a concrete case of queueing systems (the $G/M/1$ queue after perturbation of the service law [9], and the $M/G/1$ limit model for high retrial intensities (which is the classical $M/G/1$ system) after perturbation of the retrials parameter [10]). The same idea has been already investigated for an approximation analysis of the classical $G/G/1$ queue when the general law of service is unknown and must be estimated by different statistical methods, pointing out particularly to the impact of those taking into account the correction of boundary effects [6], see also the recent work of [7] and [8]. For example, in the latter work [8], besides of showing the interest of combining nonparametric methods with the strong stability principle for the study of the $M/G/1$ system, we also pointed out the importance of using the Student test to provide confidence intervals for the difference between the corresponding characteristics of the two considered queueing systems for the aim of comparing them (i.e., comparison of their characteristics).

Indeed, note that in practice all model parameters are imprecisely known because they are obtained by means of statistical methods. In this sense, our contribution concerns one aspect which is of some practical interest and has not been sufficiently studied in the literature; for instance, when a distribution governing a queueing system is unknown and we resort to nonparametric methods to estimate its density function. Besides, as the strong stability method assumes that the perturbation is small, then we suppose that the arrivals law of the $G/G/1$ system is close to the exponential one with parameter λ. This permits us to consider the problem of boundary bias correction [14, 16, 25] when performing nonparametric estimation of the unknown density of the law G, since the exponential law is defined on the positive real line.

It is why we use, in this paper, the tools of nonparametric density estimation to approximate the complex $G/G/1$ system by the simpler $M/G/1$ one, on the basis of the theoretical results addressed in [3] involving the strong stability of the $M/G/1$ system. When the distribution of arrivals is general but close to the exponential distribution, it is possible to approximate the characteristics of the $G/G/1$ system by those of the $M/G/1$ one, if we prove the fact of stability (see [2]). This substitution of characteristics is not justified without a prior estimation of the corresponding approximation error. This gives rise to the following question: Is it possible to precise the error of the proximity between the two systems?

Note that unlike [5] where kernel density estimation was used for the study of the strong stability of $M/M/1$ system, we consider here two new aspects. The first one concerns the model motivation: in queueing theory, there exist explicit formulas to determine some performance measures of the $M/G/1$ system. Unfortunately, for the $G/G/1$ system, these exact formulas are not known. So, if we suppose that the $G/G/1$ system is close to the $M/G/1$ one, then we can use the formulas obtained for the $M/G/1$ system to approximate the $G/G/1$ system characteristics. The second point deals with the use of a new class of nonparametric density estimation to remove boundary effects. This is the class of flexible estimators, for instance asymmetric kernels and smoothed histograms. Note also that unlike [6] where the perturbation concerns the service duration, we perturb the arrival flux.

This article is organized as follows: In Sect. 2, we describe the considered queueing models and we present briefly the strong stability of the $M/G/1$ system. In Sect. 3, we first provide a short review of boundary bias correction techniques in nonparametric density estimation, then we give the main results of this paper which are illustrated by a numerical case study based on simulation results.

2 Approximating $G/G/1$ Queue by the $M/G/1$ One Using Strong Stability Approach

2.1 *Description of the Models*

Consider a $G/G/1$ $(FIFO, \infty)$ queueing system with general service times distribution H and general inter-arrival times probability distribution G. The following notations are used: T_n (the arrival time of the nth customer), θ_n (the departure time of the nth customer), and γ_n (the time till the arrival of the following customer after θ_n). Let us designate by $\nu_n = \nu(\theta_n + 0)$ the number of customers in the system immediately after θ_n. ξ_n represents the service time of the nth customer arriving at the system. It is proved that $X_n = (\nu_n, \gamma_n)$ forms a homogeneous Markov chain with state space $\mathbb{N} \times \mathbb{R}^+$ and transition operator $Q = (Q_{ij})_{i,j\geq 0}$, where $Q_{ij}(x, dy) = P(\nu_{n+1} = j, \gamma_{n+1} \in dy/\nu_n = i, \gamma_n = x)$ is defined by (see [3]):

$$Q_{ij} = \begin{cases} q_j(dy), & \text{if } i = 0; \\ q_{j-i}(x, dy), & \text{if } i \geq 1, \ \ j \geq i; \\ p(x, dy), & \text{if } j = i-1, \ \ i \geq 1; \\ 0, & \text{otherwise}; \end{cases} \tag{1}$$

where

$$\begin{cases} q_j(dy) = \int P(T_j \le u < T_{j+1}, T_{j+1} - u \in dy)dH(u); \\ q_j(x, dy) = \int\limits_x^{\infty} P(T_j \le u - x < T_{j+1}, T_{j+1} - (u - x) \in dy)dH(u); \\ p(x, dy) = \int\limits_0^{x} P(x - u \in dy)dH(u). \end{cases}$$

Let us also consider an $M/G/1$ $(FIFO, \infty)$ system with exponential inter-arrivals distribution, E_λ, with parameter λ and take the same service times distribution than the $G/G/1$ one. We introduce the corresponding following notations: $\bar{T}_n$, $\bar{\theta}_n$, $\bar{\gamma}_n$, $\bar{\nu}_n = \bar{\nu}(\bar{\theta}_n - 0)$ and ξ_n defined as above. The transition operator $\bar{Q} = \left(\bar{Q}_{ij}\right)_{i,j>0}$ of the corresponding Markov chain $\bar{X}_n$ in the $M/G/1$ system has the same form as in (1), where

$$\begin{cases} \overline{q}_j(dy) = p_j E_\lambda(dy), \;\; \overline{q}_j(x, dy) = p_j(x)E_\lambda(dy), \;\; \overline{p}(x, dy) = p(x, dy); \\ p_j = \int \exp(-\lambda u)\frac{(\lambda u)^j}{j!}dH(u); \\ p_j(x) = \int_x^\infty \exp(-\lambda(u-x))\frac{(\lambda(u-x))^j}{j!}dH(u). \end{cases}$$

Let us suppose that the arrival flow of the $G/G/1$ system is close to the Poisson one. This proximity is then characterized by the metric:

$$w^* = w^*(G, E_\lambda) = \int \varphi^*(t)|G - E_\lambda|(dt), \tag{2}$$

where φ^* is a weight function and $|a|$ designates the variation of the measure a. We take $\varphi^*(t) = e^{\delta t}$, with $\delta > 0$. In addition, we use the following notations:

$$\begin{cases} E^* = \int \varphi^*(t)E_\lambda(dt), \\ G^* = \int \varphi^*(t)G(dt), \end{cases}$$

$$w_0 = w_0(G, E_\lambda) = \int |G - E_\lambda|(dt). \tag{3}$$

2.2 Strong Stability Criterion

For a general framework on the strong stability method, the reader is referred to [2, 19]. However, it is interesting to recall the following basic definition.

Definition 1 (See [2, 19]) The Markov chain X with transition kernel P and invariant measure π is said to be υ-strongly stable with respect to the norm $\|.\|_\upsilon$ (defined for each measure α as follows: $\|\alpha\|_\upsilon = \sum\limits_{j\ge 0} \upsilon(j)|\alpha_j|$), if $\|P\|_\upsilon < \infty$ and each stochastic kernel Q in some neighborhood $\{Q : \|Q - P\|_v < \epsilon\}$ has a unique invariant measure $\mu = \mu(Q)$ and $\|\pi - \mu\|_v \to 0$ as $\|Q - P\|_v \to 0$.

2.3 Strong Stability Bounds

The following theorem determines the v-strong stability conditions of the $M/G/1$ system after a small perturbation of the arrivals law. It also gives the estimates of the deviations of both the transition kernels and the stationary distributions.

Theorem 1 ([3]) *Suppose that in the $M/G/1$ system, the following ergodicity condition holds:*

(a) $\lambda\, \mathbf{E}(\xi) < 1$; *(b)* $\exists a > 0 : \quad \mathbf{E}(e^{a\xi}) = \int e^{au} dH(u) < \infty.$

Suppose also that $E^* < \infty$ and $\beta_0 = \sup(\beta : H^*(\lambda - \lambda\beta) < \beta)$, where H^* is the Laplace transform of the probability density of the service times. Then, for all β such that $1 < \beta < \beta_0$, the Markov chain $\overline{X}_n$ is v-strongly stable for the function $v(n, t) = \beta^n[\exp(-\alpha t) + c^{-1}\varphi^*(t)]$, where:

$$\alpha > 0, \quad c = \frac{\beta E^*}{1 - \rho}, \quad \text{and} \quad \rho = \frac{H^*(\lambda - \lambda\beta) + \beta}{2\beta} < 1.$$

In addition, if $G^* < \infty$, and $w_0 \leq \frac{(\beta_0 - \beta)}{\beta_0^2}$, then we have the margin between the transition operators:

$$\|Q - \overline{Q}\|_v \leq w^*(1 + \beta) + w_0 G^*(1 + \lambda\beta)\frac{\beta_0^4}{(\beta_0 - \beta)^2}.$$

Moreover, if the general distribution of arrivals G is such that:

$$w^*(G, E_\lambda) \leq \frac{1 - \rho}{2c_0(1 + c)}(1 + \beta + c_1)^{-1},$$

$$w_0(G, E_\lambda) \leq \frac{(\beta_0 - \beta)}{\beta_0^2},$$

we obtain the deviation between the stationary distributions π and $\bar{\pi}$ associated, respectively, to the Markov chains X_n and $\bar{X}_n$, given by:

$$Er := \|\pi - \overline{\pi}\| \leq 2[(1 + \beta)w^* + c_1 w_0]c_0 c_2(1 + c), \tag{4}$$

where $c_0, \ c_1, \ c_2$ are defined as follows:

$$c_0' \leq c_0,$$

where

$$c_0^{'} = 1 + \frac{(1-\lambda m)(\beta-1)(2-\rho)E^*}{2(1-\rho)^2} \text{ and } m = \mathbf{E}(\xi),$$

$$c_1 = G^*(1+\lambda\beta)\frac{\beta_0^4}{(\beta_0-\beta)^2},$$

$$c_2 = \frac{(1-\lambda m)(\beta-1)(2-\rho)}{2(1-\rho)\beta}.$$

Note that the bound in formula (4) of Theorem 1 involves the computation of w^* and w_0 and methods to do so will be discussed in the following.

3 Nonparametric Estimation for Approximating the $G/G/1$ System by the $M/G/1$ One

We want to apply nonparametric density estimation methods to determine the variation distances w_0 and w^* defined, respectively, in (2) and (3), together with the proximity error Er defined in (4) between the stationary distributions of the $G/G/1$ and $M/G/1$ systems. We first give an overview of nonparametric estimation methods which are required to compute w_0 and w^* measures, then we perform a simulation study.

3.1 Nonparametric Density Estimation Methods

The most known and used nonparametric estimation method is the kernel density estimation. If $X_1, \ldots, X_n$ is a sample coming from a random variable X with probability density function f and distribution F, then the Parzen–Rosenblatt kernel estimator [21, 24] of the density $f(x)$ for each point $x \in \mathbb{R}$ is given by:

$$f_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right), \tag{5}$$

where K is a symmetric density function called the kernel and h_n is the bandwidth.

The classical symmetric kernel estimate works well when estimating densities with unbounded support. However, when these latter are defined on the positive real line $[0, \infty[$, without correction, the kernel estimates suffer from boundary effects since they have a boundary bias (the expected value of the standard kernel estimate at $x = 0$ converges to the half value of the underlying density when f is twice continuously differentiable on its support $[0, +\infty)$ [14, 25]). In fact, using a fixed symmetric kernel is not appropriate for fitting densities with bounded supports as a weight is given outside the support.

Several approaches for handling the boundary effects in nonparametric density estimation have been introduced. They propose the use of estimators based on flexible kernels (asymmetric kernels [14, 16] and smoothed histograms [14]). They are very simple in implementation, free of boundary bias, always nonnegative, their support matches the support of the probability density function to be estimated, and their rate of convergence for the mean integrated squared error is $O(n^{-4/5})$.

Below, are briefly discussed the estimators which we will use in the context of this paper.

Reflection Method

Schuster [25] suggests creating the mirror image of the data on the other side of the boundary and then applying the estimator (5) for the set of the initial data and their reflection. $f(x)$ is then estimated, for $x \geq 0$, as follows:

$$\tilde{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} \left[K\left(\frac{x - X_j}{h_n}\right) + K\left(\frac{x + X_j}{h_n}\right) \right]. \tag{6}$$

Asymmetric Gamma Kernel Estimator

Asymmetric kernels [14, 16] are defined by the form

$$\hat{f}_b(x) = \frac{1}{n} \sum_{i=1}^{n} K(x, b)(X_i), \tag{7}$$

where b is the bandwidth and the asymmetric kernel K can be taken as a Gamma density K_G with the parameters $(x/b + 1, b)$ given by

$$K_G\left(\frac{x}{b} + 1, b\right)(t) = \frac{t^{x/b} e^{-t/b}}{b^{x/b+1} \Gamma(x/b + 1)}. \tag{8}$$

Smoothed Histograms

Smoothed histograms [14] are defined by the form

$$\hat{f}_k(x) = k \sum_{i=0}^{+\infty} \omega_{i,k} p_{ki}(x), \tag{9}$$

where the random weights $\omega_{i,k}$ are given by

$$\omega_{i,k} = F_n\left(\frac{i+1}{k}\right) - F_n\left(\frac{i}{k}\right),$$

where F_n is the empiric distribution, k is the smoothing parameter, and $p_{ki}(.)$ can be taken as a Poisson distribution with parameter kx,

$$p_{ki}(x) = e^{-kx}\frac{(kx)^i}{i!}, \qquad i = 0, 1, \dots \tag{10}$$

3.2 Algorithm

To realize this work, we use the discrete event simulation approach [4] to simulate the according systems and we elaborate an algorithm which follows the following steps:

(1) Generation of a sample of size n of general arrivals distribution G with theoretical density $g(x)$.
(2) Use of a nonparametric estimation method to estimate the theoretical density function $g(x)$ by a function denoted in general $g_n^*(x)$.
(3) Calculation of the mean arrival rate given by:
$\lambda = 1/\int xdG(x) = 1/\int xg(x)dx = 1/\int xg_n^*(x)dx$.
(4) Verification, in this case, of the strong stability conditions given in Sect. 2.3. For calculation considerations, the variation distances w_0 and w^* are given, respectively, by: $w_0 = \int |G - E_\lambda|(dx) = \int |g_n^* - e_\lambda|(x)dx$ and $w^* = \int e^{\delta x}|G - E_\lambda|(dx) = \int e^{\delta x}|g_n^* - e_\lambda|(x)dx$, where $\delta > 0$.
(5) Computation of the minimal error on the stationary distributions of the considered systems according to (4).

Simulation studies were performed under Matlab 7.1 environment. The Epanechnikov kernel [26] is used throughout for estimators involving symmetric kernels. The bandwidth h_n is chosen to minimize the criterion of the "least squares cross-validation" [18]. The smoothing parameters b and k are chosen according to a bandwidth selection method which leads to an asymptotically optimal window in the sense of minimizing L_1 distance [14].

3.3 Numerical Example

We consider a $G/G/1$ system such that the general inter-arrivals distribution G is assumed to be a Gamma distribution with parameters $\alpha = 0.7, \beta = 2$, denoted $\Gamma(0.7, 2)$, with a theoretical density $g(x)$ and the service times distribution is Cox2 with parameters: $\mu_1 = 3$, $\mu_2 = 10$, $a = 0.005$.

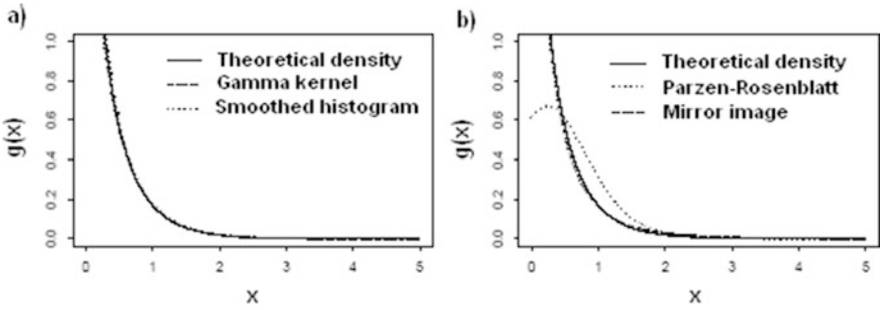

Fig. 1 Theoretical density $g(x) = \gamma(0.7, 2)(x)$ and estimated densities. (**a**) Gamma kernel and smoothed histogram estimates; (**b**) Parzen-Rosenblatt and Mirror image estimates. Taken from [8].

Table 1 Performance measures with different estimators

	$g(x)$	$g_n(x)$	$\tilde{g}_n(x)$	$\hat{g}_b(x)$	$\hat{g}_k(x)$
Mean arrival rate λ	1.6874	1.5392	1.6503	1.6851	1.6840
Traffic intensity of the system $\frac{\lambda}{\mu}$	0.1562	0.1578	0.1570	0.1564	0.1567
Variation distance w_0	0.0096	0.1287	0.0114	0.0102	0.0105
Variation distance w^*	0.0183	0.2536	0.0311	0.0206	0.0224
Error on stationary distributions Er	0.0356		0.0452	0.0378	0.0377

Taken from [8].

By generating a sample coming from the $\Gamma(0.7, 2)$ distribution, we use the different nonparametric estimators given, respectively, in (5)–(10) to estimate the theoretical density $g(x)$.

For these estimators, we take the sample size $n = 200$ and the number of simulations $R = 100$.

Curves of the theoretical and estimated densities are illustrated in Fig. 1 (taken from [8]). Different performance measures are listed in Table 1 (taken from [8]).

Interpretation of Results

Figure 1 shows that the use of nonparametric density estimation methods taking into account the correction of boundary effects improves the quality of the estimation (compared to the curve of the Parzen–Rosenblatt estimator, those of mirror image, asymmetric Gamma kernel, and smoothed histogram estimators are closer to the curve of the theoretical density).

We note in Table 1 that the approximation error on the stationary distributions of the $G/G/1$ and $M/G/1$ systems was given when applying nonparametric density estimation methods by considering the correction of boundary effects such in the cases of using the mirror image estimator ($Er = 0.0452$), asymmetric Gamma kernel estimator ($Er = 0.0378$), and smoothed histogram ($Er = 0.0377$). In addition, these two last errors are close to the one given when using the theoretical density $g(x)$ ($Er = 0.0356$). But, when applying the Parzen–Rosenblatt estimator which does not take into account the correction of boundary effects, the approximation error Er on the stationary distributions of the quoted systems could not be given. This shows the importance of the smallness of the proximity error of the two corresponding arrival distributions of the considered systems, characterized by the variation distances w_0 and w^*.

4 Conclusion

We use statistical techniques, for instance, nonparametric density estimation with boundary effects considerations to measure the performance of the strong stability method in a $M/G/1$ queueing system after perturbation of the arrival flow.

The obtained results show particularly the interest of nonparametric estimation methods and the techniques of correction of boundary effects to determine the approximation error of the stationary distributions between two queueing systems when applying the strong stability method in order to substitute the characteristics of a complex real system by another simpler ideal one.

Note that, in practice, all model parameters are imprecisely known because they are obtained by means of statistical methods. That is why the strong stability inequalities will allow us to numerically estimate the uncertainty shown during this analysis. In our case, if one had real data, then one could apply the kernel density method to estimate the density function. By combining the techniques of correction of boundary effects with the calculation of the variation distance characterizing the proximity of the quoted systems, one will be able to check if this density is sufficiently close to that of the Poisson law (or that of the exponential law), and apply then the strong stability method to approximate the characteristics of the real system by those of a classical one.

A close field of some practical interest is networks of queues. Indeed, for modeling some complex physical systems, a simple queue is not sufficient, so we may resort to networks of queues. However, few among them have simple analytic solutions. This is mainly due to the difficulty of studying the properties of inter-stations fluxes. In fact, the only known exact results are those of networks having the product form property, such as the Jackson networks. There comes the interest of analyzing such networks by combining the strong stability aspect and the boundary correction techniques.

References

1. Abbas, K., Aïssani, D.: Strong stability of the embedded Markov chain in an $GI/M/1$ queue with negative customers. Appl. Math. Model. **34**(10), 2806–2812 (2010)
2. Aïssani, D., Kartashov, N.V.: Ergodicity and stability of Markov chains with respect to operator topology in the space of transition kernels. Doklady Akademii Nauk Ukrainskoi SSR (Seriya A) **11**, 3–5 (1983)

3. Aïssani, D., Kartashov, N.V.: Strong stability of the imbedded Markov chain in an $M/G/1$ system. Theory Probab. Math. Stat. **29**, 1–5 (1984)
4. Banks, J., Carson, J.S., Nelson, B.L.: Discrete-Event System Simulation. Prentice Hall, Upper Saddle River (1996)
5. Bareche, A., Aïssani, D.: Kernel density in the study of the strong stability of the M/M/1 queueing system. Oper. Res. Lett. **36**(5), 535–538 (2008)
6. Bareche, A., Aïssani, D.: Statistical techniques for a numerical evaluation of the proximity of $G/G/1$ and $G/M/1$ queueing systems. Comput. Math. Appl. **61**(5), 1296–1304 (2011)
7. Bareche, A., Aïssani, D.: Combinaison des méthodes de stabilité forte et d'estimation non paramétrique pour l'approximation de la file d'attente $G/G/1$. J. européen des systèmes automatisés **47**(1-2-3), 155–164 (2013)
8. Bareche, A., Aïssani, D.: Statistical methodology for approximating $G/G/1$ queues by the strong stability technique. In: Proceedings of the 3rd International Conference on Operations Research and Enterprise Systems, pp. 241–248 (2014). doi:10.5220/0004834002410248
9. Benaouicha, M., Aïssani, D.: Strong stability in a $G/M/1$ queueing system. Theory Probab. Math. Stat. **71**, 25–36 (2005)
10. Berdjoudj, L., Aïssani, D.: Strong stability in retrial queues. Theory Probab. Math. Stat. **68**, 11–17 (2003)
11. Berdjoudj, L., Benaouicha, M., Aïssani, D.: Measure of performances of the strong stability method. Math. Comput. Model. **56**, 241–6246 (2012)
12. Bouallouche, L., Aïssani, D.: Measurement and performance of the strong stability method. Theory Probab. Math. Stat. **72**, 1–9 (2006)
13. Bouallouche, L., Aïssani, D.: Performance analysis approximation in a queueing system of type $M/G/1$. Math. Meth. Oper. Res. **63**(2), 341–356 (2006)
14. Bouezmarni, T., Scaillet, O.: Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data. Econ. Theory **21**, 390–412 (2005)
15. Boukir, L., Bouallouche-Medjkoune, L., Aïssani, D.: Strong stability of the batch arrival queueing systems. Stoch. Anal. Appl. **28**(1), 8–25 (2010)
16. Chen, S.X.: Probability density function estimation using gamma kernels. Ann. Inst. Stat. Math. **52**, 471–480 (2000)
17. Hamadouche, N., Aïssani, D.: Approximation in the $M_2/G_2/1$ queue with preemptive priority. Methodol. Comput. Appl. Probab. **13**(3), 563–581 (2010)
18. Jones, M.C., Marron, J.S., Sheather, S.J.: A brief survey of bandwidth selection for density estimation. J. Am. Stat. Assoc. **91**, 401–407 (1996)
19. Kartashov, N.V.: Strong Stable Markov Chains. TbiMC Scientific Publishers, VSPV, Utrecht (1996)
20. Kleinrock, L.: Queueing Systems, vols. 1 and 2. Wiley, New York (1976)
21. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**, 1065–1076 (1962)
22. Rabta, B., Aïssani, D.: Strong stability in an (R, s, S) inventory model. Int. J. Prod. Econ. **4**, 159–171 (2005)
23. Rahmoune, F., Aïssani, D.: Strong stability of queue with multi vacations of the server. Stoch. Anal. Appl. **26**(3), 665–678 (2008)
24. Rosenblatt, M.: Remarks on some non-parametric estimates of a density function. Ann. Math. Stat. **27**, 832–837 (1956)
25. Schuster, E.F.: Incorporating support constraints into nonparametric estimation of densities. Commun. Stat. Theory Meth. **14**, 1123–1136 (1985)
26. Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)

Stochastic Analysis of an M/G/1 Retrial Queue with FCFS

Mohamed Boualem, Mouloud Cherfaoui, Natalia Djellab, and Djamil Aïssani

Abstract The main goal of this paper is to investigate stochastic analysis of a single server retrial queue with a First-Come-First-Served (FCFS) orbit and non-exponential retrial times using the monotonicity and comparability methods. We establish various results for the comparison and monotonicity of the underlying embedded Markov chain when the parameters vary. Moreover, we prove stochastic inequalities for the stationary distribution and some simple bounds for the mean characteristics of the system. We validate stochastic comparison method by presenting some numerical results illustrating the interest of the approach.

1 Introduction

Queueing systems with repeated attempts have been widely used to model many problems in telecommunication and computer systems [1, 4, 19]. The essential feature of a retrial queue is that arriving customers who find all servers busy are obliged to abandon the service area and join a retrial group, called orbit, in order to try their luck again after some random time. For a detailed review of the main results

M. Boualem (✉)
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Technology, University of Bejaia, Bejaia 06000, Algeria
e-mail: robertt15dz@yahoo.fr

M. Cherfaoui
Research Unit LaMOS (Modeling and Optimization of Systems), University of Bejaia, Bejaia 06000, Algeria

Department of Mathematics, University of Biskra, 07000 Biskra, Algeria
e-mail: mouloudcherfaoui2013@gmail.com

N. Djellab
Laboratory LaPS, Department of Mathematics, University of Annaba, Annaba 23000, Algeria
e-mail: djellab@yahoo.fr

D. Aïssani
Research Unit LaMOS (Modeling and Optimization of Systems), Faculty of Exact Sciences, University of Bejaia, Bejaia 06000, Algeria
e-mail: lamos_bejaia@hotmail.com

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*, Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_8

and the literature on this topic the reader is referred to the monographs [2, 11]. In recent years, there has been an increasing interest in the investigation of the retrial phenomenon in cellular mobile network, see [3, 10, 15, 16, 24] and the references therein, and in many other telecommunication systems including star-like local area networks [14], wavelength-routed optical networks [26], circuit-switched systems with hybrid fiber-coax architecture [13], wireless sensor networks [25], etc.

It is well known that for the retrial queues we need to establish how the customers in orbit access to the server. The time between successive repeated attempts is important in telephony, where a call receiving a busy signal does not wait the termination of the busy condition. The most usual protocol described in the classical theory of retrial queues is the so-called classical retrial policy in which each source in orbit repeats its call after an exponentially distributed time with parameter θ. So, there is a probability $n\theta dt + o(dt)$ of a new retrial in the next interval $(t,\ t + dt)$ given that n customers are in orbit at time t. Such a policy has been motivated by applications in modeling subscriber's behavior in telephone networks since the 1940s. In past years, technology has considerably evolved. The literature on retrial queues describes several retrial protocols specific to some modern computer and communication networks in which the time between two successive repeated attempts is controlled by an electronic device and consequently, is independent of the number of units applying for service. In this case, the probability of a repeated attempt during $(t,\ t+dt)$, given the orbit is not empty, is $(1-\delta_{0,n})\alpha\ dt+o(dt)$ where $\delta_{0,n}$ denotes Kronecker's delta and n is the number of repeated customers. This type of retrial discipline is called the constant retrial policy.

An examination of the literature on the retrial queues reveals the remarkable fact that the non-homogeneity caused by the flow of repeated attempts is the key to understand most analytical difficulties arising in the study of retrial queues. Many efforts have been devoted to deriving performance measures such as queue length, waiting time, busy period distributions, and so on. However, these performance characteristics have been provided through transform methods which have made the expressions cumbersome and the obtained results cannot be put into practice. In the last decade there has been a trend towards the research of approximations and bounds. Qualitative properties of stochastic models constitute a basic theoretical basis for approximation methods. Some important approaches are monotonicity and comparability which can be investigated using the stochastic comparison method based on the general theory of stochastic orderings. Stochastic orders represent an important tool for many problems in probability and statistics [18, 20–23].

Stochastic comparison is a mathematical tool used in the performance study of systems modeled by continuous or discrete-time Markov chains. The general idea of this method is to bound a complex system by a new system, easier to solve and providing performance measures bounds. Many papers treat stochastic comparison methods of queueing systems with repeated attempts. Boualem et al. [6] investigate some monotonicity properties of an $M/G/1$ queue with constant retrial policy in which the server operates under a general exhaustive service and multiple vacation policy relative to strong stochastic ordering and convex ordering. These results imply in particular simple insensitive bounds for the stationary queue length distribution. Boualem et al. [7] use the tools of a qualitative analysis to investigate various monotonicity properties for an $M/G/1$ retrial queue with classical retrial policy and Bernoulli feedback. The obtained results allow to place in a prominent position the insensitive bounds for both the stationary distribution and the conditional distribution of the stationary queue of the considered model. Mokdad and Castel-Taleb [17] propose to use a mathematical method based on stochastic comparisons of Markov chains in order to derive bounds on performance indices of fixed and mobile networks. Their main objective consists in finding Markovian bounding models with reduced state spaces, which are easier to solve. They apply the methodology to performance evaluation of complex telecommunication systems modeled by large size Markov chains which cannot be solved by exact methods. They propose to define intuitively bounding systems in order to compute bounds on performance measures. Using stochastic comparison methods, they prove that the new systems represent bounds for the exact ones. To validate their approach and illustrate its interest, they present some numerical results. Bušić and Fourneau [9] illustrate through examples how monotonicity may help for performance evaluation of mobile networks, by considering two different applications. In the first application, they assume that a Markov chain of the model depends on a parameter that can be estimated only up to a certain level and they have only an interval that contains the exact value of the parameter. Instead of taking an approximated value for the unknown parameter, they show how monotonicity properties of the Markov chain can be used to take into account the error bound from the measurements. In the second application, they consider a well-known approximation method: the decomposition into Markovian submodels. They show that the monotonicity property may help to derive bounds for Markovian submodels and are sufficient conditions for convergence of iterative algorithms which are often designed to give approximations. More recently, Boualem et al. [8] investigate various monotonicity properties of a single server retrial queue with general retrial times using the mathematical method based on stochastic comparisons of Markov chains in order to derive bounds on performance indices. Bounds are derived for the mean characteristics of the busy period, number of customers served during a busy period, number of orbit busy periods, and waiting times. Boualem [5] addresses monotonicity properties of the single server retrial queue with no waiting room and server subject to active breakdowns, that is, the service station can fail only during the service period. The obtained results give insensitive bounds for the stationary distribution of the considered embedded Markov chain related to the model in the study. Numerical illustrations are provided to support the results.

In this paper we consider an $M/G/1$ retrial queue with non-exponential retrial times under the special assumption that only the customer at the head of the orbit queue is allowed to occupy the server. The performance characteristics of such a system are available in the literature (see [12]). The author obtains relevant performance characteristics expressed in terms of generating functions and Laplace transforms. However, there still remains the issue that numerical inversion is required for actually computing numbers and derive useable results. Indeed, it is sometimes possible to obtain the generating function and/or Laplace transforms of an unknown probability distribution but not to invert the generating function or the Laplace transforms to obtain an explicit form of the distribution. Moreover, the error for numerical inversion is difficult to control. For example, if we compare two systems which are “close” then it might be that due to the numerical error in the inversion, we may take the wrong system to perform better. Based on the relevant performance characteristics obtained by Gómez-Corral [12], we consider in our paper a qualitative analysis which is another field of own right to establish insensitive bounds on some performance measures by using the stochastic analysis approach relative to the theory of stochastic orderings . Finally, the effects of various parameters on the performance of the system have been examined numerically.

This paper is arranged as follows. In the next section, we describe the considered mathematical model. In Sect. 3, we introduce some pertinent definitions and notions of the three most important orderings. Section 4 focusses on monotonicity of the transition operator and gives comparability conditions of two transition operators. Stochastic inequalities for the stationary number of customers in the system are discussed in Sect. 5. The last section is devoted to the practical applications.

2 Mathematical Model

Primary customers arrive in a Poisson process with rate λ. If the server is free, the primary customer is served immediately and leaves the system after service completion. Otherwise, the customer leaves the service area and enters the retrial group in accordance with an FCFS discipline. We assume that only the customer at the head of the orbit is allowed for access to the server. If the server is busy upon retrial, the customer joins the orbit again. Such a process is repeated until the customer finds the server idle and gets the requested service at the time of a retrial. Successive inter-retrial times of any customer follow an arbitrary law with common probability distribution function $A(x)$, Laplace-Stieltjes transform $\mathbb{L}_A(s)$ and first moment α_1. The service times are independently and identically distributed with probability distribution function $B(x)$, Laplace-Stieltjes transform $\mathbb{L}_B(s)$ and first two moments β_1, β_2. We suppose that inter-arrival times, retrial times, and service times are mutually independent.

The main characteristic of this queue is that, at any service completion, a competition between an exponential law and a general retrial time distribution determines the next customer who accesses the service facility. Thus, the retrial discipline does not depend on the orbit length.

Let τ_n be the time of the n-th departure and Z_n the number of customers in the orbit just after the time τ_n. We have the following fundamental recursive equation:

$$Z_{n+1} = Z_n + v^{n+1} - \delta_{Z_{n+1}},$$

where v^{n+1} is the number of primary customers arriving at the system during the service time which ends at τ_{n+1}. Its distribution is given by:

$$b_j = \mathbb{P}(v^{n+1} = j) = \int_0^\infty (\lambda x)^j (j!)^{-1} e^{-\lambda x} dB(x),\ j \geq 0,$$

with generating function $b(z) = \sum_{j\geq 0} b_j z^j = \mathbb{L}_B(\lambda(1-z))$.

The Bernoulli random variable $\delta_{Z_{n+1}}$ is equal to 1 or 0 depending on whether the customer who leaves the system at time τ_{n+1} proceeds from the orbit or otherwise.

The sequence of random variables $\{Z_n, n \geq 1\}$ forms an embedded Markov chain for our queueing system which is irreducible and aperiodic on the state-space $\mathbb{N}$. The stability condition is given in [12] as follows: $\rho < \mathbb{L}_A(\lambda)$, where $\rho = \lambda\beta_1$ is the load of the system.

3 Stochastic Orders

Stochastic orders are useful in comparing random variables measuring certain characteristics in many areas. Such areas include insurance, operations research, queueing theory, survival analysis, and reliability theory (see [22]). The simplest comparison is through comparing the expected value of the two comparable random variables. First, we define some notions on stochastic ordering which will be used in the context of the paper. For more details see [20–23].

Definition 1 Let $F(x)$ and $G(x)$ be two distribution functions of nonnegative random variables X and Y, respectively. Then:

(a) $F \leq_{\text{st}} G$ iff $F(x) \geq G(x)$ or $\overline{F}(x) = 1 - F(x) \leq \overline{G}(x)$, $\forall x \geq 0$.
(b) $F \leq_{icx} G$ iff $\int_x^{+\infty} \overline{F}(u)d(u) \leq \int_x^{+\infty} \overline{G}(u)d(u)$, $\forall x \geq 0$.
(c) $F \leq_L G$ iff $\int_0^{+\infty} \exp(-sx)dF(x) \geq \int_0^{+\infty} \exp(-sx)dG(x)$, $\forall s \geq 0$.

Definition 2 If the random variables of interest are of discrete type and $\alpha = (\alpha_n)_{n\geq 0}$, $\beta = (\beta_n)_{n\geq 0}$ are the corresponding distributions, then the above definitions can be given in the following form:

(a) $\alpha \leq_{\text{st}} \beta$ iff $\overline{\alpha}_m = \sum_{n\geq m} \alpha_n \leq \overline{\beta}_m = \sum_{n\geq m} \beta_n$, for all m.
(b) $\alpha \leq_{icx} \beta$ iff $\overline{\overline{\alpha}}_m = \sum_{n\geq m}\sum_{k\geq n} \alpha_k \leq \overline{\overline{\beta}}_m = \sum_{n\geq m}\sum_{k\geq n} \beta_k$, for all m.
(c) $\alpha \leq_L \beta$ iff $\sum_{n\geq 0} \alpha_n z^n \geq \sum_{n\geq 0} \beta_n z^n$, for all $z \in [0, 1]$.

Definition 3 Let X be a positive random variable with distribution function F:

1. F is *HNBUE* (Harmonically New Better than Used in Expectation) iff $F \leq_{icx} F^*$,
2. F is of class $\mathscr{L}$ iff $F \geq_L F^*$,

where F^* is the exponential distribution function with the same mean as F.

The ageing classes are linked by the inclusion chain:

$$\text{NBU (New Better than Used)} \subset \text{NBUE (New Better than Used in Expectation)} \subset \text{HNBUE} \subset \mathscr{L}.$$

4 Monotonicity and Comparability of the Transition Operator

The one-step transition probabilities of $\{Z_n, n \geq 1\}$ are defined by

$$p_{nm} = \begin{cases} (1 - \mathbb{L}_A(\lambda))b_{m-n} + \mathbb{L}_A(\lambda)b_{m-n+1}, & \text{for } n \neq 0 \text{ and } m \geq 0, \\ b_m, & \text{for } n = 0 \text{ and } m \geq 0. \end{cases} \tag{1}$$

Let Θ be the transition operator of an embedded Markov chain which associates to every distribution $\alpha = \{\alpha_m\}_{m\geq 0}$ a distribution $\Theta\alpha = \{\beta_m\}_{m\geq 0}$ such that

$$\beta_m = \sum_{n\geq 0} \alpha_n p_{nm}.$$

Theorem 1 *The operator Θ is monotone with respect to the orders $\leq_{st}$ and $\leq_{icx}$.*

Proof The operator Θ is monotone with respect to $\leq_{\text{st}}$ if and only if $\overline{p}_{n-1m} \leq \overline{p}_{nm}$, and is monotone with respect to $\leq_{icx}$ if and only if $2\overline{\overline{p}}_{nm} \leq \overline{\overline{p}}_{n-1m} + \overline{\overline{p}}_{n+1m}$ for all $n,\ m$, where

$\overline{p}_{nm} = \sum_{l\geq m} p_{nl}$ and $\overline{\overline{p}}_{nm} = \sum_{k\geq m} \overline{p}_{nk} = \sum_{k\geq m}\sum_{l\geq k} p_{nl}$.

In our case:

$$\overline{p}_{nm} - \overline{p}_{n-1m} = (1 - \mathbb{L}_A(\lambda))b_{m-n} + \mathbb{L}_A(\lambda)b_{m-n+1} > 0.$$
$$\overline{\overline{p}}_{n-1m} + \overline{\overline{p}}_{n+1m} - 2\overline{\overline{p}}_{nm} = (1 - \mathbb{L}_A(\lambda))b_{m-n-1} + \mathbb{L}_A(\lambda)b_{m-n} > 0.$$

□

Theorems 2 till 4, we give comparability conditions of two transition operators. Consider two $M/G/1$ retrial queues with non-exponential retrial times with parameters $\lambda^{(i)}, A^{(i)}, B^{(i)}$. Let Θ^i be the transition operator of the embedded Markov chain, in the i-th system, $i = 1, 2$.

Theorem 2 *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_{st} B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $\Theta^1 \leq_{st} \Theta^2$, i.e., for any distribution α, we have $\Theta^1\alpha \leq_{st} \Theta^2\alpha$.*

Proof From Stoyan [23], it is well known that to prove $\Theta^1 \leq_{\rm st} \Theta^2$, we have to show the following inequality:

$$\overline{p}_{nm}^{(1)} \leq \overline{p}_{nm}^{(2)}, \ \forall \ n, m.$$

We have

$$\overline{p}_{nm}^{(1)} = (1 - \mathbb{L}_{A^{(1)}}(\lambda^{(1)}))b_{m-n}^{(1)} + \overline{b}_{m-n+1}^{(1)}.$$

Since $\lambda^{(1)} \leq \lambda^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then

$$\mathbb{L}_{A^{(1)}}(\lambda^{(1)}) \geq \mathbb{L}_{A^{(2)}}(\lambda^{(2)}),$$

and

$$\overline{p}_{nm}^{(1)} \leq (1 - \mathbb{L}_{A^{(2)}}(\lambda^{(2)}))b_{m-n}^{(1)} + \overline{b}_{m-n+1}^{(1)}.$$

But

$$(1 - \mathbb{L}_{A^{(2)}}(\lambda^{(2)}))b_{m-n}^{(1)} + \overline{b}_{m-n+1}^{(1)} = (1 - \mathbb{L}_{A^{(2)}}(\lambda^{(2)}))\overline{b}_{m-n}^{(1)} + \mathbb{L}_{A^{(2)}}(\lambda^{(2)})\overline{b}_{m-n+1}^{(1)}.$$

Using these inequalities we get:

$$\overline{p}_{nm}^{(1)} \leq (1 - \mathbb{L}_{A^{(2)}}(\lambda^{(2)}))\overline{b}_{m-n}^{(2)} + \mathbb{L}_{A^{(2)}}(\lambda^{(2)})\overline{b}_{m-n+1}^{(2)} = \overline{p}_{nm}^{(2)}.$$

□

Theorem 3 *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_{icx} B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $\Theta^1 \leq_{icx} \Theta^2$.*

Proof The proof is similar to that of Theorem 2. □

Theorem 4 *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_L B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $\Theta^1 \leq_L \Theta^2$.*

Proof Let α be a distribution and $\Theta_\alpha = \beta$, where

$$\beta_m = \sum_{n\geq 0} \alpha_n p_{nm} = \alpha_0 b_m + \sum_{n\geq 1} \alpha_n p_{nm}, \ \text{for all } m \geq 0.$$

The generating function of β is given by

$$G(z) = \sum_{m\geq 0} \beta_m z^m = \alpha_0 b(z) + \frac{1}{z} b(z)(\alpha(z) - \alpha_0)(z + (1 - z)\mathbb{L}_A(\lambda)).$$

If the conditions of Theorem 4 are fulfilled, then

$$b^{(1)}(z) \geq b^{(2)}(z) \text{ and } (1-z)\mathbb{L}_{A^{(1)}}(\lambda^{(1)}) \geq (1-z)\mathbb{L}_{A^{(2)}}(\lambda^{(2)}), \ \forall z \in [0,1].$$

Hence $G^{(1)}(z) \geq G^{(2)}(z)$. □

5 Stochastic Inequalities for the Stationary Distribution

Consider two $M/G/1$ retrial queues with non-exponential retrial times. Let $\pi_n^{(1)}$, $\pi_n^{(2)}$ be the corresponding stationary distributions of the number of customers in the system.

Theorem 5 *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_s B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $\{\pi_n^{(1)}\} \leq_s \{\pi_n^{(2)}\}$, where $\leq_s$ represents one of the symbols $\leq_{\text{st}}$ or $\leq_{icx}$.*

Proof Using Theorems 1–3 which state that Θ^i are monotone with respect to the order $\leq_s$ and $\Theta^1 \leq_s \Theta^2$, we have by induction $\Theta^{1,n}\alpha \leq_s \Theta^{2,n}\alpha$ for any distribution α, where $\Theta^{i,n} = \Theta^i(\Theta^{i,n-1}\alpha)$. Taking the limit, we obtain the stated result. Indeed, $\Theta^1\alpha_n^1 = \mathbb{P}[Z_k^1 = n] \leq_s \mathbb{P}[Z_k^2 = n] = \Theta^2\alpha_n^2$, when $k \to \infty$, we have $\{\pi_n^{(1)}\} \leq_s \{\pi_n^{(2)}\}$. □

Theorem 6 *If in the $M/G/1$ retrial queue with general retrial times the service time distribution $B(x)$ is HNBUE (Harmonically New Better than Used in Expectation) and the retrial time distribution is $\mathscr{L}$, then $\{\pi_n\} \leq_{icx} \{\pi_n^*\}$, where $\{\pi_n^*\}$ is the stationary distribution of the number of customers in the $M/M/1$ retrial queue with exponential retrial with the same parameters.*

Proof Consider an auxiliary $M/M/1$ retrial queue with exponentially distributed retrial time $A^*(x)$ and service time $B^*(x)$. If $B(x)$ is *HNBUE* and $A(x)$ is $\mathscr{L}$, then $B(x) \leq_{icx} B^*(x)$ and $A(x) \leq_L A^*(x)$. Therefore, by using Theorem 5, we deduce the statement of this theorem. □

6 Practical Aspect

Assume that we have two $M/G/1$ retrial queues with non-exponential retrial times with parameters $\lambda^{(1)}$, $A^{(1)}$, $B^{(1)}$ and $\lambda^{(2)}$, $A^{(2)}$, $B^{(2)}$, respectively. Let $L^{(i)}$, $I^{(i)}$, $N_b^{(i)}$ and $W^{(i)}$ be the busy period length, the number of customers served during a busy period, the number of orbit busy periods which take place in $]0, L^{(i)}]$ and the waiting time , respectively, in the i-th system, $i = 1, 2$.

Theorem 7 *If $\lambda^{(1)} \leq \lambda^{(2)}$, $B^{(1)} \leq_s B^{(2)}$ and $A^{(1)} \leq_L A^{(2)}$, then $\mathbb{E}(L^{(1)}) \leq \mathbb{E}(L^{(2)})$ and $\mathbb{E}(I^{(1)}) \leq \mathbb{E}(I^{(2)})$, where $\leq_s$ is one of the symbols $\leq_{st}$, $\leq_{icx}$, $\leq_L$.*

Proof Gómez-Corral [12] shows that

$$\mathbb{E}(L) = \frac{\beta_1}{\mathbb{L}_A(\lambda) - \lambda\beta_1} \text{ and } \mathbb{E}(I) = \frac{\mathbb{L}_A(\lambda)}{\mathbb{L}_A(\lambda) - \lambda\beta_1},$$

which are increasing with respect to λ and β_1, decreasing with respect to $\mathbb{L}_A(.)$. Under conditions of Theorem 7, we obtain the desired inequalities. □

Theorem 8 *For any $M/G/1$ retrial queue,*

$$\mathbb{E}(L) \le \mathbb{E}(L)_{\text{Upper}} = \frac{\beta_1}{e^{-\lambda\alpha_1} - \lambda\beta_1},$$

$$\mathbb{E}(I) \le \mathbb{E}(I)_{\text{Upper}} = \frac{e^{-\lambda\alpha_1}}{e^{-\lambda\alpha_1} - \lambda\beta_1}.$$

If A and B are of class $\mathscr{L}$, then

$$\mathbb{E}(L) \ge \mathbb{E}(L)_{\text{Lower}} = \frac{\beta_1(1+\lambda\alpha_1)}{1-\lambda\beta_1(1+\lambda\alpha_1)},$$

$$\mathbb{E}(I) \ge \mathbb{E}(I)_{\text{Lower}} = \frac{1}{1-\lambda\beta_1(1+\lambda\alpha_1)}.$$

Proof We consider auxiliary $M/D/1$ and $M/M/1$ retrial queues with the same arrival rates λ, mean service times β_1 and mean retrial times α_1. A represents Dirac distribution at α_1 for the $M/D/1$ system, and represents the exponential distribution for the $M/M/1$ system. Using the theorem above we obtain the stated results. □

Theorem 9 *If $\lambda^{(1)} \le \lambda^{(2)}$, $B^{(1)} \le_{st} B^{(2)}$ and $A^{(1)} \le_L A^{(2)}$, then $\mathbb{E}(N_b^{(1)}) \le \mathbb{E}(N_b^{(2)})$ and $\mathbb{E}(W^{(1)}) \le \mathbb{E}(W^{(2)})$.*

Proof Gómez-Corral [12] shows that

$$\mathbb{E}(N_b) = \frac{1-\mathbb{L}_B(\lambda)}{\mathbb{L}_B(\lambda)} \text{ and } \mathbb{E}(W) = \frac{\lambda\beta_2 + 2\beta_1(1-\mathbb{L}_A(\lambda))}{2(\mathbb{L}_A(\lambda) - \lambda\beta_1)}.$$

These quantities are increasing with respect to λ, β_1 and β_2, decreasing with respect to $\mathbb{L}_B(.)$ and $\mathbb{L}_A(.)$. Under the conditions of Theorem 9, we obtain the desired inequalities. □

Theorem 10 *For any $M/G/1$ retrial queue,*

$$\mathbb{E}(N_b) \le \mathbb{E}(N_b)_{\text{Upper}} = e^{\lambda\beta_1} - 1,$$

$$\mathbb{E}(W) \le \mathbb{E}(W)_{\text{Upper}} = \frac{\lambda\beta_2 + 2\beta_1(1-e^{-\lambda\alpha_1})}{2(e^{-\lambda\alpha_1} - \lambda\beta_1)}.$$

If A and B are of class $\mathscr{L}$, then

$$\mathbb{E}(N_b) \geq \mathbb{E}(N_b)_{\text{Lower}} = \lambda\beta_1,$$
$$\mathbb{E}(W) \geq \mathbb{E}(W)_{\text{Lower}} = \frac{\lambda\beta_2(1+\lambda\alpha_1)+2\lambda\beta_1\alpha_1}{2(1-\lambda\beta_1(1+\lambda\alpha_1))}.$$

Proof The proof is similar to that of Theorem 8. □

6.1 Numerical Application

We give a numerical illustration concerning the mean busy period length $\mathbb{E}(L)_{A(x)}$ and the mean waiting time $\mathbb{E}(W)_{A(x)}$ in the $M/M/1$ retrial queue with general retrial times given respectively in Theorems 8 and 10. To this end, for the retrial time distributions $A(x)$, we have considered the most representative distributions which are:

1. Exponential (exp): $A(x) = 1 - e^{-\alpha_1 x}$.
2. Two-Stage Erlang (E_2): $A(x) = 1 - (1 - 2\gamma x)e^{-2\gamma x}$.
3. Gamma (Γ): $A(x) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} e^{-t/b} dt$.
4. Two-Stage Hyper-Exponential (H_2): $A(x) = 1 - pe^{-\gamma_1 x} - (1-p)e^{-\gamma_2 x}$.

In Table 1 we present the values of the system parameters according the above cases.

The obtained results are presented in Figs. 1 and 2. From these results, we note that:

- The lower bound $\mathbb{E}(L)_{\text{Lower}}$ (respectively, $\mathbb{E}(W)_{\text{Lower}}$) is nothing else than the mean length of the busy period $\mathbb{E}(L)$ (respectively, the mean waiting time $\mathbb{E}(W)$) in the $M/M/1$ retrial queue with exponential retrial times.
- The inequality $\mathbb{E}(L)_{A(x)} \leq \mathbb{E}(L)_{\text{Upper}}$ (respectively, $\mathbb{E}(W)_{A(x)} \leq \mathbb{E}(W)_{\text{Upper}}$) always holds. In addition, if the law $A \in \mathscr{L}$, then the inequality $\mathbb{E}(L)_{\text{Lower}} \leq \mathbb{E}(L)_{A(x)}$ (respectively, $\mathbb{E}(W)_{\text{Lower}} \leq \mathbb{E}(W)_{A(x)}$) holds.
- If α_1 and ρ are small enough then the mean length of the busy period (respectively, the mean waiting time) in the system is closer to the $\mathbb{E}(L)_{A(x)}$ (respectively, $\mathbb{E}(W)$), in other words, closer to the $\mathbb{E}(L)_{\text{Lower}}$ (respectively, $\mathbb{E}(W)_{\text{Lower}}$).

Table 1 Different values of the system parameters

ρ	λ	β_1	α_1	γ	(a, b)	(p, γ_1, γ_2)
0.3		0.3	[0.500, 0.400, 0.333, 0.286, 0.250]		$a = 3.5$	$p = 0.3$
0.6	1	0.6	[0.125, 0.143, 0.167, 0.200, 0.250]	$\frac{2}{\alpha_1}$		$\gamma_1 = 4$
0.8		0.8	[0.083, 0.091, 0.100, 0.111, 0.125]		$b = \frac{\alpha_1}{3.5}$	$\gamma_2 = \frac{(1-p)\alpha_1\gamma_1}{(\gamma_1 - \alpha p)}$

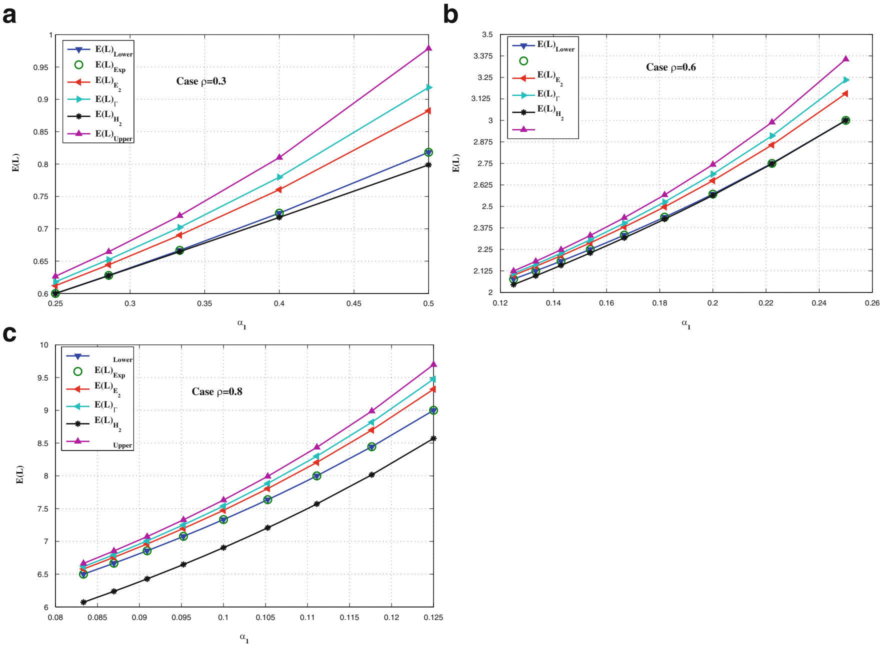

Fig. 1 Comparison of the $\mathbb{E}(L)$ in $M/M/1$ queue with general retrial times versus α_1

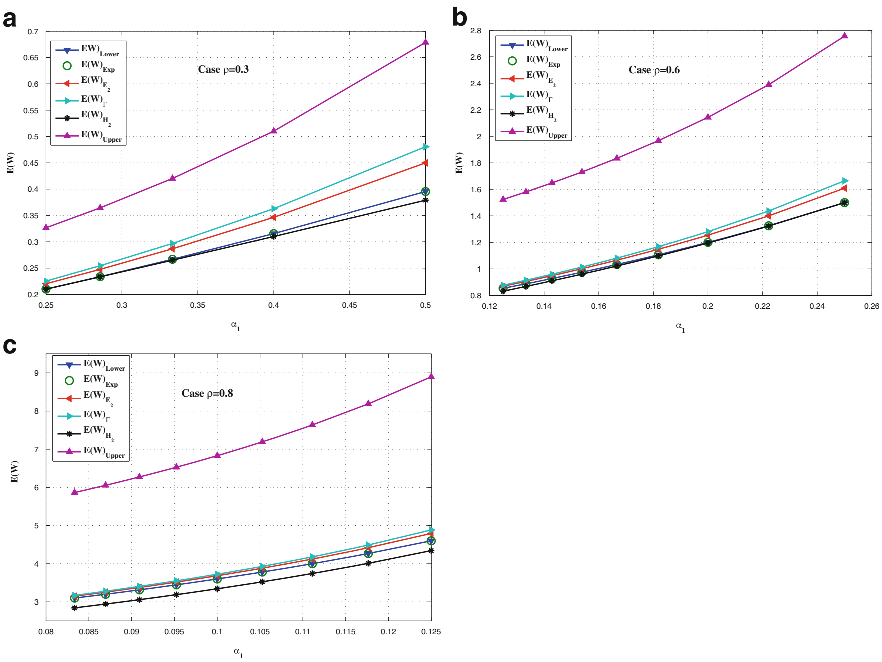

Fig. 2 Comparison of the $\mathbb{E}(W)$ in $M/M/1$ queue with general retrial times versus α_1

- If the distribution of the retrial time is close to the exponential distribution in the Laplace transform, then the exact value $\mathbb{E}(L)_{A(x)}$ (respectively, $\mathbb{E}(W)_{A(x)}$) is closer to the lower bound $\mathbb{E}(L)_{\text{Lower}}$ (respectively, $\mathbb{E}(W)_{\text{Lower}}$) (see the case of $E(L)_{E_2}$ and $\mathbb{E}(W)_{E_2}$).
- Both considered characteristics depend closely on the inter-retrial times distribution and its first moment α_1. In addition, this dependence appears clearly in the case of heavy traffic , i.e., when $\rho \to 1$.

7 Conclusion

The main result of this paper consists to give insensitive bounds for the stationary distribution and some performance measures of the considered embedded Markov chain by using the theory of stochastic orderings. The result is confirmed by numerical illustrations.

In conclusion, the monotonicity approach holds promise for the solution of several systems with repeated attempts. Hence, it is worth noting that our approach can be further extended to more complex systems.

References

1. Artalejo, J.R., Gómez-Corral, A.: Channel idle periods in computer and telecommunication systems with customer retrials. Telecommun. Syst. **24**, 29–46 (2003)
2. Artalejo, J.R., Gómez-Corral, A.: Retrial Queueing Systems: A Computational Approach. Springer, Berlin (2008)
3. Artalejo, J.R., López-Herrero, M. J.: Cellular mobile networks with repeated calls operating in random environment. Comput. Oper. Res. **37**, 1158–1166 (2010)
4. Artalejo, J.R., Pla, V.: On the impact of customer balking, impatience and retrials in telecommunication systems. Comput. Math. Appl. **57**, 217–229 (2009)
5. Boualem, M.: Insensitive bounds for the stationary distribution of a single server retrial queue with server subject to active breakdowns. Adv. Oper. Res. **2014**, Article ID 985453, 12 pp. (2014)
6. Boualem, M., Djellab, N., Aïssani, D.: Stochastic inequalities for $M/G/1$ retrial queues with vacations and constant retrial policy. Math. Comput. Model. **50**, 207–212 (2009)
7. Boualem, M., Djellab, N., Aïssani, D.: Stochastic approximations and monotonicity of a single server feedback retrial queue. Math. Probl. Eng. **2012**, Article ID 536982, 13 pp. (2012)
8. Boualem, M., Djellab, N., Aïssani, D.: Stochastic bounds for a single server queue with general retrial times. Bull. Iranian Math. Soc. **40**, 183–198 (2014)
9. Bušić, A., Fourneau, J.M.: Monotonicity and performance evaluation: applications to high speed and mobile networks. Cluster Comput. **15**, 401–414 (2012)
10. Economou, A., López-Herrero, M.J.: Performance analysis of a cellular mobile network with retrials and guard channels using waiting and first passage time measures. Eur. Trans. Telecommun. **20**, 389–401 (2009)
11. Falin, G.I., Templeton, J.G.C.: Retrial Queues. Chapman and Hall, London (1997)
12. Gómez-Corral, A.: Stochastic analysis of a single server retrial queue with the general retrial times. Nav. Res. Logist. **46**, 561–581 (1999)

13. Houck D.J., Lai, W.S.: Traffic modeling and analysis of hybrid fiber-coax systems. Comput. Netw. ISDN Syst. **30**, 821–834 (1998)
14. Janssens, G.K.: The quasi-random input queueing system with repeated attempts as a model for a collision-avoidance local area network. IEEE Trans. Commun. **45**, 360–364 (1997)
15. Liu, X., Fapojuwo, A.O.: Performance analysis of hierarchical cellular networks with queueing and user retrials. Int. J. Commun. Syst. **19**, 699–721 (2006)
16. Machihara, F., Saitoh, M.: Mobile customers model with retrials. Eur. J. Oper. Res. **189**, 1073–1087 (2008)
17. Mokdad, L., Castel-Taleb, H.: Stochastic comparisons: a methodology for the performance evaluation of fixed and mobile networks. Comput. Commun. **31**, 3894–3904 (2008)
18. Müller, A., Stoyan, D.: Comparison Methods for Stochastic Models and Risk. Wiley, New York (2002)
19. Pustova, S.V.: Investigation of call centers as retrial queuing systems. Cybern. Syst. Anal. **46**, 494–499 (2010)
20. Ross, S.M.: Stochastic Processes. Wiley, New York (1983)
21. Shaked, M., Shanthikumar, J.G.: Stochastic Orders and their Applications. Academic, San Diego (1994)
22. Shaked, M., Shanthikumar, J.G.: Stochastic Orders. Academic, New York (2006)
23. Stoyan, D.: Comparison Methods for Queues and Other Stochastic Models. Wiley, New York (1983)
24. Van Do, T.: A new computational algorithm for retrial queues to cellular mobile systems with guard channels. Comput. Ind. Eng. **59**, 865–872 (2010)
25. Wüchner, P., Sztrik, J., de Meer, H.: Modeling wireless sensor networks using finite-source retrial queues with unreliable orbit. In: Hummel, K.A. et al. (eds.) PERFORM 2010 (Haring Festschrift), LNCS 6821, pp. 73–86 (2011)
26. Xue, F., Yoo, S.J.B., Yokoyama, H., Horiuchi, Y.: Performance analysis of wavelength-routed optical networks with connection request retrials. In: IEEE International Conference on Communication ICC 2005, vol. 3, pp. 1813–1818. IEEE, New York (2005)

Dynalets: A New Tool for Biological Signal Processing

Jacques Demongeot, Ali Hamie, Olivier Hansen, and Mustapha Rachdi

Abstract The biological information coming from electro-physiologic signal sensors needs compression for an efficient medical use or for retaining only the pertinent explanatory information about the mechanisms at the origin of the recorded signal. When the signal is periodic in time and/or space, classical compression procedures like Fourier and wavelets transforms give good results concerning the compression rate, but provide in general no additional information about the interactions between the elements of the living system producing the studied signal. Here, we define a new transform called Dynalets based on Liénard differential equations susceptible to model the mechanism at the source of the signal and we propose to apply this new technique to real signals like ECG.

1 Introduction

There are different manners to represent a biological signal aiming to both (a) explain the mechanisms having produced it and (b) facilitate its use in medical applications. The biological signals come from electro-physiologic signal sensors like ECG and have to be compressed for an efficient medical use by clinicians or to retain only the pertinent explanatory information about the mechanisms at the origin of the recorded signal for the researchers in life sciences. When the signal is periodic in time and/or space, the classical compression processes like Fourier and wavelets transforms give good results concerning the compression rate, but bring in general no supplementary information about the interactions between elements of the living system producing the studied signal. Here, we define a new transform called Dynalets based on Liénard differential equations, susceptible to

J. Demongeot (✉) • A. Hamie • O. Hansen
AGIM CNRS FRE 3405, Faculty of Medicine, University J. Fourier of Grenoble, 38700 La Tronche, France
e-mail: Jacques.Demongeot@agim.eu; stat_hamie@hotmail.com; Olivier.Hansen@agim.eu

M. Rachdi
AGIM CNRS FRE 3405, UPMF, UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France
e-mail: Mustapha.Rachdi@upmf-grenoble.fr; Mustapha.Rachdi@agim.eu

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_9

model the mechanism that is the source of the signal and we propose to apply this new technique to real signals like ECG.

2 Fourier and Haley Wavelet Transforms

The Fourier transform comes from the aim by J. Fourier to represent in a simple way the functions used in physics, notably in the heat propagation (in 1807, cf. [10]). He used a base of functions made of the solutions of the simple pendulum differential equation (cf. a trajectory in Fig. 1):

$$\frac{dx}{dt} = y, \frac{dy}{dt} = -\omega^2 x, \tag{1}$$

its general solution being:

$$x(t) = k \cos \omega t,$$
$$y(t) = -k\omega \sin \omega t.$$

By using the polar coordinates θ and ρ defined from the variables x and $z = -y/\omega$, we get the new differential system:

$$\frac{d\theta}{dt} = \omega,$$
$$\frac{d\rho}{dt} = 0, \tag{2}$$

with $\theta = \arctan(z/x)$ and $\rho^2 = x^2 + z^2$.

The polar system is conservative, its Hamiltonian function being defined by:

$$H(\theta, \rho) = \omega\rho.$$

The solutions:

$$x(t) = k \cos \omega t,$$
$$z(t) = k \sin \omega t$$

have 2 degrees of freedom, k and ω, respectively the amplitude and the frequency of the signal, and they constitute an orthogonal base, when we choose for ω the multiples (called harmonics) of a fundamental frequency ω_0.

After the seminal theoretical works by Y. Meyer [12, 17], I. Daubechies [2], and S. Mallat [15], J. Haley defined a simple wavelet transform for representing signals in astrophysics (in 1997, cf. [9]). He used a base of functions made of the solutions of the damped pendulum differential equation (cf. a trajectory in Fig. 1):

$$\frac{dx}{dt} = y,$$
$$\frac{dy}{dt} = -(\omega^2 + \tau^2)x - 2\tau y,$$

its general solution being:

$$x(t) = ke^{-\tau t}\cos\omega t,$$
$$y(t) = -ke^{-\tau t}(\omega\sin\omega t + \tau\cos\omega t),$$

by using the polar coordinates θ and ρ defined from the variables x and $z = -y/\omega - \tau x/\omega$, we get the differential system:

$$\frac{d\theta}{dt} = \omega,$$
$$\frac{d\rho}{dt} = -\tau\rho.$$

This polar system is a potential function being defined by:

$$P(\theta, \rho) = -\omega\theta + \frac{\tau\rho^2}{2}.$$

The solutions:

$$x(t) = ke^{-\tau t}\cos\omega t,$$
$$z(t) = ke^{-\tau t}\sin\omega t$$

have 3 degrees of freedom, k, ω, and τ, the last parameter being the exponential time constant responsible for pendulum damping.

3 The Van der Pol System

For Dynalets transform, we propose to use a base of functions made of the solutions of the relaxation pendulum differential equation (cf. a trajectory in Fig. 1 Top), which is a particular example of the most general Liénard differential equation:

$$\frac{dx}{dt} = y,$$

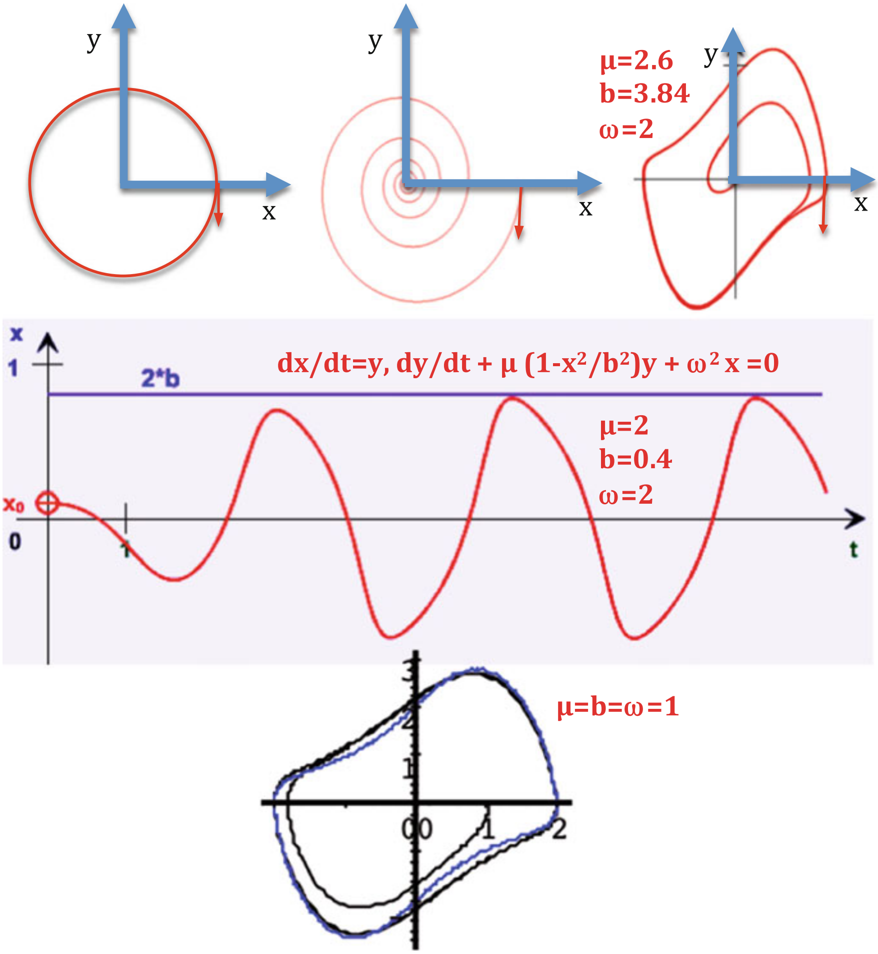

Fig. 1 Figure 1 *Top left*: a simple pendulum trajectory. *Top middle*: a damped pendulum trajectory. *Top right*: Van der Pol limit cycle. *Middle*: relaxation oscillation of Van der Pol oscillator without external forcing. *Bottom*: representation of the harmonic contour lines $H(x, y) = 2.024$

$$\frac{dy}{dt} = -R(x)x + Q(x)y,$$

which is specified in Van der Pol case by choosing:

$$R(x) = \omega^2 \text{ and } Q(x) = \mu\left(1 - \frac{x^2}{b^2}\right).$$

Its general solution is not algebraic, but approximated by a family of polynomials [4, 5, 9, 14].

The Van der Pol system is a potential-Hamiltonian system, with P and H functions (Fig. 2 Top left), H being for example approximated at order 4, when $\omega = b = 1$, by [4, 5]:

$$H(x,y) = \frac{(x^2+y^2)}{2} - \frac{\mu xy}{2} + \frac{\mu yx^3}{8} - \frac{\mu xy^3}{8},$$

which allows to obtain the equation of its limit cycle (cf. Fig. 1 Bottom): $H(x,y) \approx 2.024$. The Van der Pol system has 3 degrees of freedom, b, ω, and μ, the last an-harmonic parameter being responsible of the asymptotic stability of the pendulum limit cycle, which is symmetrical with respect to the origin, but not revolution symmetrical. These parameters receive different interpretations:

– μ appears as an-harmonic reaction term: when $\mu = 0$, the equation is that of the simple pendulum, i.e., a sine wave oscillator, whose amplitude depends on initial conditions and relaxation oscillations are observed even with small initial conditions (Figs. 1 and 2 Middle), whose period T near the bifurcation value $\mu = 0$ equals $2\pi/\mathrm{Im}\beta$; β is the eigenvalue of the Jacobian matrix J of the Van der Pol equation at the origin and for $\omega = b = 1$:

$$J = \begin{pmatrix} 0 & 1 \\ -1 & \mu \end{pmatrix},$$

whose characteristic polynomial is equal to: $\beta^2 - \mu\beta + 1 = 0$, hence:

$$\beta = \frac{\mu \pm \sqrt{\mu^2 - 4}}{2} \text{ and } T \approx 2\pi + \frac{\pi\mu^2}{4}.$$

– b looks as a term of control: when $x > b$, the derivative of y is negative, acting as a moderator on the velocity. The maximum of the oscillations amplitude of x is about $2b$ whatever initial conditions and values of the other parameters. More precisely, the amplitude $a_x(\mu)$ of x is estimated by $2b < a_x(\mu) < 2.024b$, for every $\mu > 0$, and when μ is small, $a_x(\mu)$ is estimated by:

$$a_x(\mu) \approx \frac{\left(2 + \frac{\mu^2}{6}\right)b}{1 + \frac{7\mu^2}{96}}, \text{ (cf. [9, 14]).}$$

The half-amplitude of y $a_y(\mu)$ is obtained for $dy/dt = 0$, that is approximately for $x = b$ if ωb is small. Then $a_y(\mu)$ is the dominant root of the following algebraic equation: $H(b, a_y(\mu)) = 2.024$.

– ω is a frequency parameter, when μ is small and the period is about $2\pi/\omega$. When $\mu \gg 1$, the period T of the limit cycle is determined mainly by the time during which the system stays around the states where y is $O(1/\mu)$. The oscillations

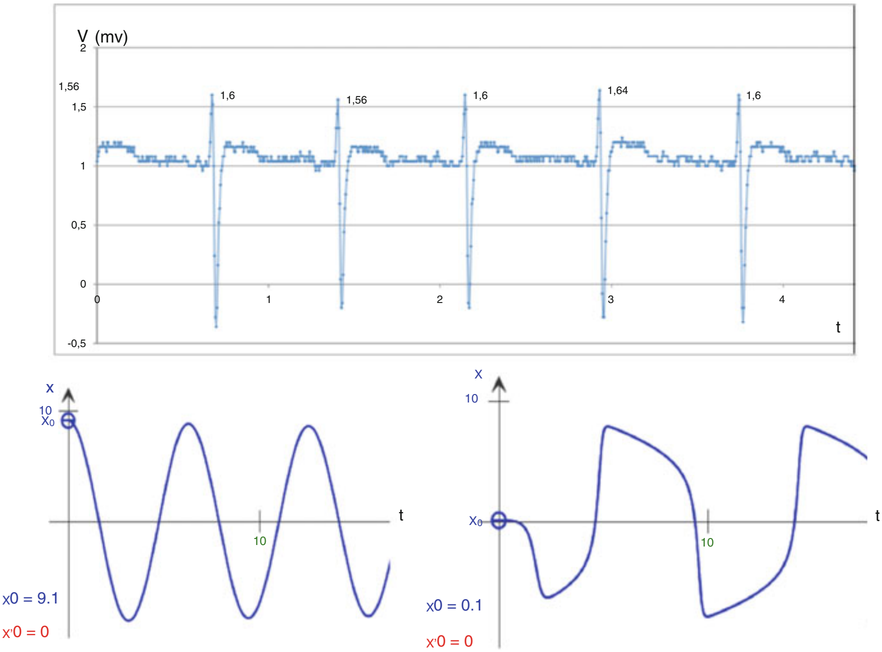

Fig. 2 *Top*: original ECG signal ($V1$ derivation). *Bottom*: representation of different waves from Van der Pol oscillator simulations, from the symmetric type (*left*, for $\mu = 0.4, b = 4, \omega = 1$) to the relaxation type (*right*, for $\mu = 4, b = 4, \omega = 1$) showing the progressive relaxation phenomenon (simulations done using [1])

period T is roughly estimated to be $T \approx \mu(3 - 2\log(2))$ (cf. [1]), and the system can be rewritten as:

$$\frac{d\chi}{dt} = \zeta,$$

$$\frac{d\zeta}{dt} = -\omega^2\chi + \mu\left(1 - \frac{\chi^2}{\mu^2}\right)\zeta \approx -\omega^2\chi + \mu\zeta,$$

with change of variables: $\chi = \mu x/b, \zeta = \mu y/b$.

4 The Dynalets Transform

The Dynalets transform consists in identifying a Liénard system-based interactions mechanism between its variables (well expressed by its Jacobian matrix) analogue to those of the experimentally studied system, whose limit cycle is the nearest (in the sense of the Δ set distance, to the signal in the phase plane (xOy), where $y = dx/dt$.

For example, the Jacobian interaction graph of the Van der Pol system contains a couple of positive and negative tangent circuits. Practically, for performing the Dynalets transform it is necessary to choose: (a) the parameter μ such as the period of the Van der Pol signal equals the mean empirical (the same value for the Van der Pol and for the signal referential, chosen as the ECG signal in Fig. 2), (b) a translation of the origin of axes, then a homothetic change of variables to match the first Van der Pol. The whole approximation procedure can be done for the ECG signal (see Fig. 3 and http://www.sciences.univnantes.fr/sites/genevieve_tulloue/Meca/Oscillateurs/vdp_phase.html, http://wikimedia.org/wikipedia/commons/7/70/ECG_12derivations and [6, 8, 13, 16, 18]) involving the following steps:

1. To perform a symmetrizing of x axis in the case of derivation V1 in order to get a signal similar the ECG V5 (cf. Fig. 2 Top) and a transformation scaling with the same homothetic coefficient the x and y axes of the ECG signal, so as to adjust them to the maximum and minimum x and y of the vdP (Van der Pol) signal.
2. To perform a translation of the origin of axes of the ECG signal by adjusting the base line to a selected phase of a vdP limit cycle of same period T (called pitch period) as the ECG period.
3. To finish the approximation matching the ECG points set to the vdP limit-cycle, by minimizing the difference set distance Δ between the interiors of the ECG and vdP cycles (denoted respectively ECG and VDP, with interiors ECG_o and VDP_o) in the phase plane: $\Delta(\mathrm{ECG}_o, \mathrm{VDP}_o) = \mathrm{Area}[(\mathrm{ECG}_o \setminus \mathrm{VDP}_o) \cup (\mathrm{VDP}_o \setminus \mathrm{ECG}_o)]$, by using a Monte-Carlo method for estimating the area of the sets interior to the linear approximation of the ECG and vdP cycles, calculated from point samples $\{Xi\}_{i=1,100}$ and $\{Yi\}_{i=1,100}$.
4. To repeat the procedure for getting after the fundamental, the successive harmonics (cf. Fig. 3).

The reconstruction with the fundamental and the first harmonics gives for the Δ set distance a relative error of 8 % in the example of Fig. 3, and the reconstruction of the second harmonics just allows passing under the 5 % threshold, even by considering 3 cycles instead 4 (cf. Fig. 4).

5 The Problem of the Baseline

The Dynalet approximation [6] (and more generally the Tailored to the Problem Specificity Mathematical Transforms, or TPSMT transforms [11]) implies to be efficient the identification and removal of the baseline and we propose two new methods, one based on the expectile regression and the other on the Levy time distribution, alternative to the classical approaches of baseline filtering [3, 7, 19–23].

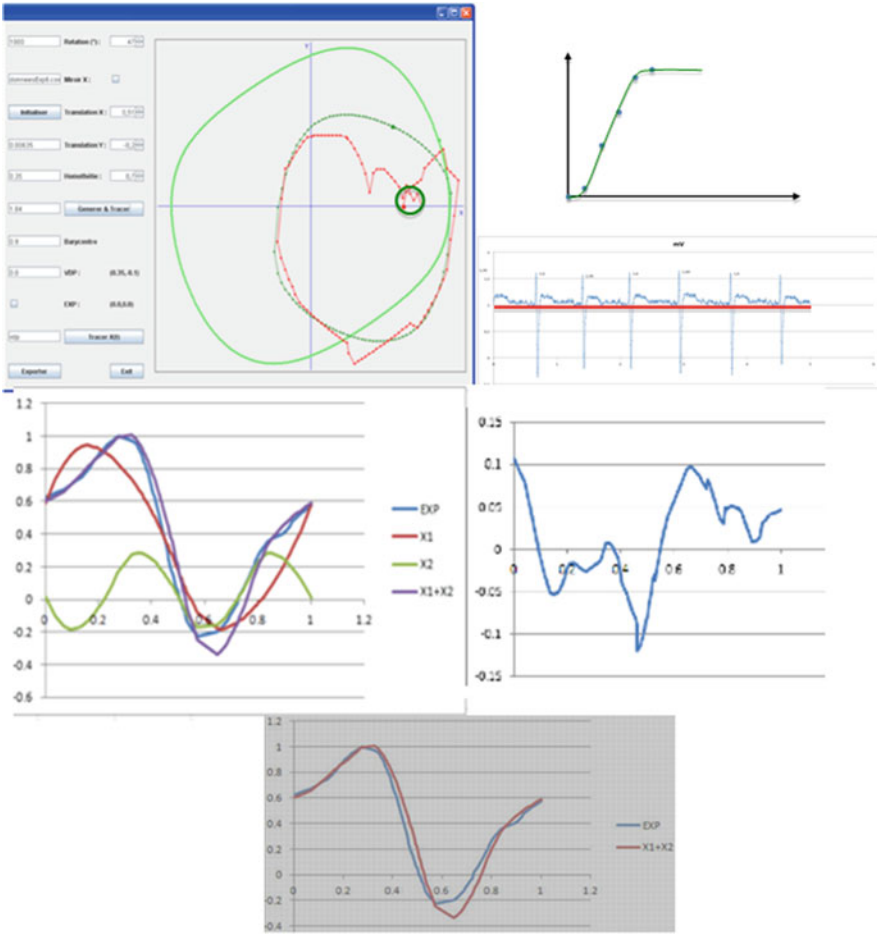

Fig. 3 *Top left*: Initial position in the phase plane *xOy* of the Van der Pol limit cycle (in *green* clear) and ECG signal (in *red*) and final fit between Van der Pol (in *dark green*) and EEG signal after transformation on *X* and *Y* axes (translation and scaling). *Top right*: experimental ECG with the evolution of the Lévy time $\lambda(\epsilon)$ corresponding to the duration of the signal passed between 0 and ϵ and the baseline obtained by averaging the signal under the threshold corresponding to the plateau value of the Léy time curve. *Middle left*: extraction of the fundamental component $X1$ (in *red*) and of the first harmonic $X2$ (in *green*) from the original experimental ECG signal (in *blue*). *Middle right*: start of the calculation of the second harmonics (in *blue*) by subtracting the fundamental plus the first harmonic component $X1 + X2$ (in *violet* on the *Middle left*) from the sampled original ECG signal. *Bottom*: comparison between the original ECG signal (in *blue*) to the reconstructed signal (in *red*) made of the fundamental plus the first harmonic (Color figure online)

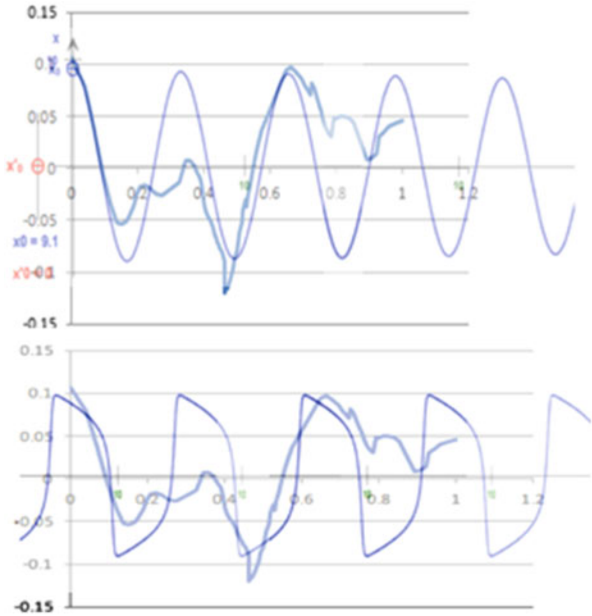

Fig. 4 *Top*: continuation of the calculation of the second harmonics (in *dark blue*) by comparing the subtracted signal of Fig. 3 to a Van der Pol signal of period $T/4$. *Bottom*: comparison between the subtracted signal of Fig. 3 to a Van der Pol signal of period $T/3$ (Color figure online)

6 Conclusion

Generalizing compression tools like Fourier or wavelets transforms is possible, if we consider that non symmetrical biological signals are often produced by mechanisms based on interactions of regulon type (i.e., possessing at least one couple of positive and negative tangent circuits inside their Jacobian interaction graph). In this case, we can replace the differential systems giving birth to biological signals by a Liénard-type equation, like the Van der Pol system classically used to model relaxation waves. The corresponding new transform, called Dynalets transform, has been built in the same spirit as the wavelets transform.

Acknowledgements We indebted to Campus France CMCU for supporting us with the grant PHC Maghreb SCIM.

References

1. Bub, G., Glass, L., Shrier, A.: Coupling dependent wavefront stability in heterogeneous cardiac cell cultures. Biophys. J. **84**, 408 (2003)
2. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. **41**, 909–996 (1988)
3. de Rooi, J.J., Eilers, P.H.C.: Mixture models for baseline estimation. Chemom. Intell. Lab. Syst. **117**, 56–60 (2012)

4. Demongeot J., Glade, N., Forest, L.: Liénard systems and potential-Hamiltonian decomposition I methodology. C. R. Math. **344**, 121–126 (2007)
5. Demongeot, J., Glade, N., Forest, L.: Liénard systems and potential-Hamiltonian decomposition II algorithm. C. R. Math. **344**, 191–194 (2007)
6. Demongeot, J., Hansen, O., Hamie, A.: Dynalets: a new tool for biological signal processing. In: Medicon'13, IFBME Proceedings, vol. 41, pp. 1250–1253. Springer, New York (2014)
7. Eilers, P.H.C., Boelens, H.F.M.: Baseline correction with asymmetric least squares smoothing. Leiden University Medical Centre Report (2005)
8. Fenton, F.H., Cherry, E.M.: Models of cardiac cell. Scholarpedia **3**, 1868 (2008)
9. Fisher, E.: The period and amplitude of the van der Pol limit cycle. J. Appl. Phys. **25**, 273–274 (1954)
10. Fourier, J.: Propagation de la Chaleur dans les corps solides. Nouv. Bull. Sci. Soc. Philomathique de Paris **6**, 112–116 (1808)
11. Glaria, A., Zepeda, H., Chabert, S., Hidalgo, S.M., Demongeot. J., Taramasco, C.: Complex adaptive systems with inference learning emergent property to estimate tailored to the problem specificity mathematical transforms: three study cases. In: ECCS'13, Complex Systems Society, Barcelona, pp. 127–129 (2013)
12. Lemarié, P.G., Meyer, Y.: Ondelettes et bases hilbertiennes. Rev. Mat. Iberoam. **2**, 1–18 (1986)
13. Lind, R., Brenner, M., Haley, S.M.: Estimation of modal parameters using a wavelet-based approach. NASA Report TM-97-206300 (1997)
14. Lopez, J.L., Abbasbandy, S., Lopez-Ruiz, R.: Formulas for the amplitude of the van der Pol limit cycle. Schol Res. Exch. **2009**, 854060 (2009)
15. Mallat, S.: A theory of multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)
16. McAllister, R.E., Noble, D., Tsien, R.W.: Reconstruction of the electrical activity of cardiac Purkinje fibres. J. Physiol. **251**, 1–59 (1975)
17. Meyer, Y.: Wavelets and operators. In: Analysis at Urbana. Lecture Notes Series, vol. 137, pp. 256–365. London Mathematical Society, London (1989)
18. Noble, D.: A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pacemaker potential. J. Physiol. **160**, 317–352 (1962)
19. Pandit, S.V.: ECG baseline drift removal through STFT. In: Proceedings of IEEE EMBS 1996, pp. 1405–1406 (1996)
20. Paul, B., Mythili, P.: ECG noise removal using GA tuned sign-data least mean square algorithm. In: IEEE ICACCT, pp. 100–103 (2012)
21. Schnabel, S.K., Eilers, P.H.C.: Optimal expectile smoothing. Comput. Stat. Data Anal. **53**, 4168–4177 (2009)
22. Sornmo, L.: Time varying digital filtering of ECG baseline wander. In: Proceedings of IEEE EMBS 1993, pp. 503–508 (1993)
23. von Borries, R.F., Pierluissi, J.H., Nazeran, H.: Wavelet transform based ECG baseline drift removal for body surface potential mapping. In: Proceedings of IEEE EMBS 2005, pp. 3891–3894 (2005)

Demography in Epidemics Modelling: The Copula Approach

Jacques Demongeot, Mohamad Ghassani, Hana Hazgui, and Mustapha Rachdi

Abstract Classical models of epidemics by Ross and McKendrick have to be revisited in order to take into account the demography (fecundity and migration) both of host and vector populations and also diffusion and mutation of infectious agents. We will study three models along different age classes of human, with and without mosquitoes by using the copula function, and we will conduct a simulation study for two of these models.

1 Introduction

The practical use of epidemic models must rely heavily on the realism put into the models. This doesn't mean that a reasonable model can include all possible effects but rather incorporate the mechanisms in the simplest possible fashion so as to maintain major components that influence disease propagation. Great care should be taken before epidemic models are used for prediction of real phenomena. However, even simple models should, and often do, pose important questions about the underlying mechanisms of infection spread and possible means of control of the endemic or epidemic. We will study as example a human epidemic disease transmitted by vectors like mosquitoes and also by humans. Both mosquitoes and human are supposed to be hosts of an infectious agent. We will take into account the demography and immunity of the human. The immunologic, genetic as well as demographic evolution of the vector and agent will be neglected.

J. Demongeot (✉) • M. Ghassani • H. Hazgui
Univ. Grenoble-Alpes, AGIM CNRS FRE 3405, Faculty of Medicine, University J. Fourier of Grenoble, 38700 La Tronche, France
e-mail: Jacques.Demongeot@agim.eu; ghassani.mohamad@yahoo.fr; Hana.Hazgui@agim.eu

M. Rachdi
Univ. Grenoble-Alpes, AGIM CNRS FRE 3405, UPMF, UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France
e-mail: Mustapha.Rachdi@upmf-grenoble.fr; Mustapha.Rachdi@agim.eu

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*, Contributions to Statistics, DOI 10.1007/978-3-319-22476-3_10

2 Model 1: Model with Three Age Classes

By introducing age classes in the classical Ross–McKendrick epidemic model [1–6], we add new demographic parameters as the fecundity rate (f_i for the susceptible humans and f_i' for the infected humans), which is equal to the mean number of off-springs a person in class i is sending to the class 1 between times t and $t+dt$, and the relative (i.e., accounting for the disease) survival (resp. death) rate b_i (resp. μ_i) equal to the probability to survive from age i to age $i+1$ (resp. to die at age i) between times t and $t+dt$. The equations of the Ross–McKendrick model [7] corresponding to three age classes with two compartments of mosquitoes (cf. Fig. 1) are:

$$\frac{\partial S_m}{\partial t} = -\frac{\zeta_1 S_m I_2}{N_H} + jI_m$$

$$\frac{\partial I_m}{\partial t} = \frac{\zeta_1 S_m I_2}{N_H} - jI_m$$

$$\begin{aligned}\frac{\partial S_1}{\partial t} &= (\gamma_1 - \beta_{11})\,E_1 - \beta_{12} S_1 E_2 - (b_1 + \mu_1)(1 - \beta_{11} - \beta_{12})S_1 + f_1 S_2 + f_1' \gamma_2 E_2 \\ &\quad + f_1'' K_2 I_2 - \frac{\zeta_2 S_1 I_m}{N_H}\end{aligned}$$

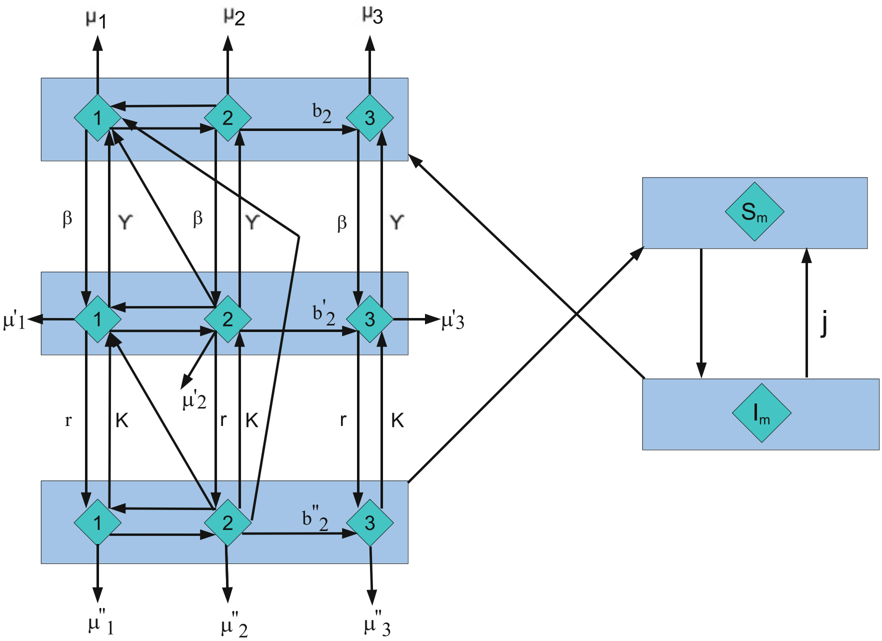

Fig. 1 Interaction graph of the model with three age classes for each of the three populations of susceptible S (*top*), infected not infectious E (*middle*) and infectious I (*bottom*) and two compartments of mosquitoes

$$
\begin{aligned}
\frac{\partial S_2}{\partial t} &= b_1(1-\beta_{11}-\beta_{12})S_1 + (\gamma_2 - \beta_{22}S_2)E_2 - \beta_{21}S_2E_1 - (\mu_2 + b_2) \\
&\quad (1-\beta_{22}-\beta_{21}-\beta_{23})S_2 - \beta_{23}S_2E_3 - \frac{\zeta_2 S_2 I_m}{N_H} \\
\frac{\partial S_3}{\partial t} &= b_2(1-\beta_{22}-\beta_{21}-\beta_{23})S_2 + (\gamma_3 - \beta_{33}S_3)E_3 - \mu_3 S_3 - \beta_{32}S_3E_2 + \gamma_2 E_2 S_3 \\
&\quad -\frac{\zeta_2 S_3 I_m}{N_H} \\
\frac{\partial E_1}{\partial t} &= (\beta_{11}E_1 + \beta_{12}E_2)S_1 - \gamma_1 E_1 + f_1'(1-\gamma_2)E_2 - (b_1' + \mu_1')(1-\gamma_1)E_1 - r_1E_1 \\
&\quad + K_1I_1 - r_{12}E_1I_2 + f_1'' K_2 I_2 + \frac{\zeta_2 S_1 I_m}{N_H} \qquad (1) \\
\frac{\partial E_2}{\partial t} &= (\beta_{21}E_1 + \beta_{22}E_2)S_2 - \gamma_2 E_2 - (\mu_2' + b_2')(1-\gamma_2)E_2 - r_2E_2 + b_1'(1-\gamma_1)E_1 \\
&\quad + K_2I_2 - r_{23}E_2I_3 + \frac{\zeta_2 S_2 I_m}{N_H} \\
\frac{\partial E_3}{\partial t} &= b_2'(1-\gamma_2)E_2 + \beta_{33}S_3E_3 - (\mu_3' + \gamma_3 + r_3)E_3 + K_3I_3 + K_{23}I_2E_3 + \beta_{23}S_2E_3 \\
&\quad + \frac{\zeta_2 S_3 I_m}{N_H} \\
\frac{\partial I_1}{\partial t} &= r_1E_1 - K_1I_1 + f_1''(1-K_2)I_2 - (b_1'' + \mu_1'')(1-K_1)I_1 - K_{12}I_1E_2 \\
\frac{\partial I_2}{\partial t} &= r_2E_2 + b_1''(1-K_1)I_1 - K_2I_2 - (b_2'' + \mu_2'')(1-K_2)I_2 - K_{23}I_2E_3 \\
\frac{\partial I_3}{\partial t} &= b_2''(1-K_2)I_2 + r_3E_3 + K_{23}E_2I_3 - (K_3 + \mu_3'')I_3
\end{aligned}
$$

where the Ss are the sizes of the three age classes for the susceptible humans, the Es their analogs for the infected humans, and Is for infectious humans. S_m (resp. I_m) denotes the number of susceptible (resp. infectious) mosquitos.

3 Model 2: Model with Two Age Classes

By taking the model (1) and removing one class of ages and the corresponding passage speeds, then we find a model with only two age classes as in Fig. 2. Equations of this model are as given in the following system, in which the f, l and h parameters represent with index 1 the fecundity and with index 2 the

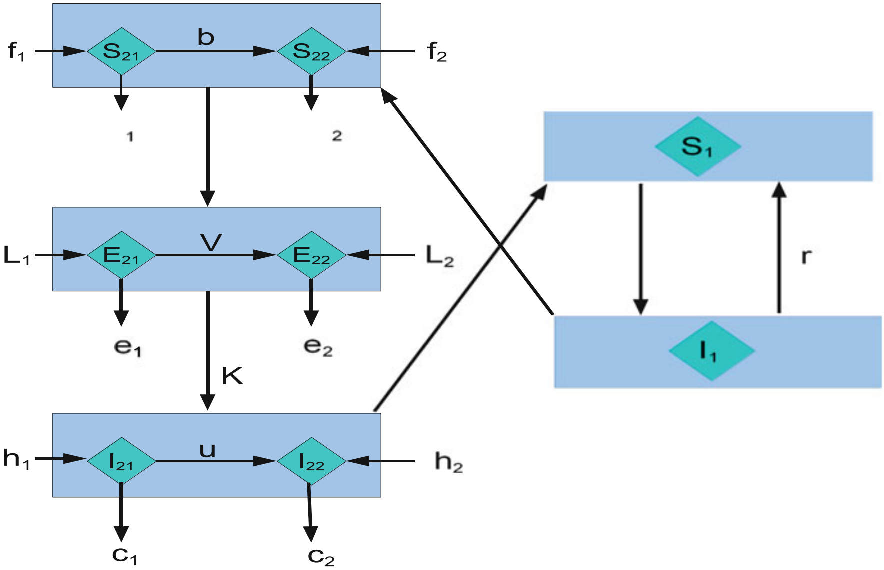

Fig. 2 Interaction graph of the model with two age classes and two compartments of mosquitoes

immigration/emigration balanve:

$$
\begin{aligned}
\frac{\partial S_1}{\partial t} &= \frac{-\beta_1 S_1 I_{22}}{N_H} + rI_1 \\
\frac{\partial I_1}{\partial t} &= \frac{\beta_1 S_1 I_{22}}{N_H} - rI_1 \\
\frac{\partial S_{21}}{\partial t} &= f_1 S_{21} - \frac{\beta_2 S_{21} I_1}{N_H} - \delta_1 S_{21} - bS_{21} \\
\frac{\partial S_{22}}{\partial t} &= f_2 S_{22} + bS_{21} - \frac{\beta_2 S_{22} I_1}{N_H} - \delta_2 S_{22} \\
\frac{\partial I_{21}}{\partial t} &= h_1 I_{21} - uI_{21} - c_1 I_{21} + KE_{21} \\
\frac{\partial I_{22}}{\partial t} &= uI_{21} + h_2 I_{22} + KE_{22} - c_2 I_{22} \\
\frac{\partial E_{21}}{\partial t} &= l_1 E_{21} - vE_{21} + \frac{\beta_2 S_{21} I_1}{N_H} - KE_{21} - e_1 E_{21} \\
\frac{\partial E_{22}}{\partial t} &= vE_{21} + l_2 E_{22} + \frac{\beta_2 S_{22} I_1}{N_H} - KE_{22} - e_2 E_{22}
\end{aligned}
\tag{2}
$$

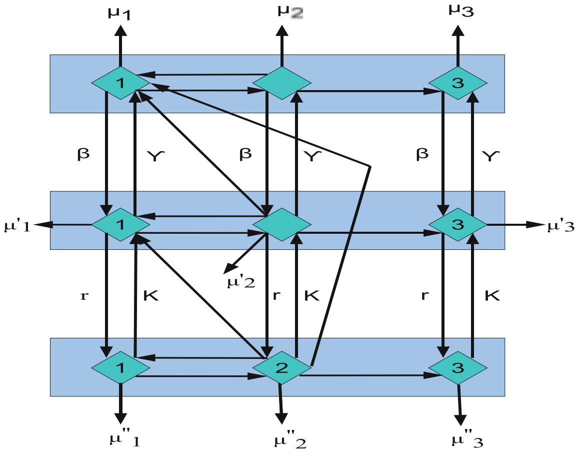

Fig. 3 Interaction graph of the model with three age classes and without mosquitoes, for the susceptible (*top*), infected not infectious (*middle*) and infectious (*bottom*) populations

4 Model 3: Three Age Classes Without Mosquitoes

By taking the model 1 and by eliminating the compartments of mosquitoes, we find the model illustrated in Fig. 3. The equations corresponding to this model are as follows:

$$
\begin{aligned}
\frac{\partial S_1}{\partial t} &= (\gamma_1 - \beta_{11})\,E_1 - \beta_{12}S_1E_2 - (b_1 + \mu_1)(1 - \beta_{11} - \beta_{12})S_1 + f_1S_2 + f_1'\gamma_2E_2 \\
&\quad + f_1''K_2I_2 \\
\frac{\partial S_2}{\partial t} &= b_1(1 - \beta_{11} - \beta_{12})S_1 + (\gamma_2 - \beta_{22}S_2)E_2 - \beta_{21}S_2E_1 - (\mu_2 + b_2)(1 - \beta_{22} \\
&\quad - \beta_{21} - \beta_{23})S_2 - \beta_{23}S_2E_3 \\
\frac{\partial S_3}{\partial t} &= b_2(1 - \beta_{22} - \beta_{21} - \beta_{23})S_2 + (\gamma_3 - \beta_{33}S_3)E_3 - \mu_3S_3 - \beta_{32}S_3E_2 + \gamma_2E_2S_3 \\
\frac{\partial E_1}{\partial t} &= (\beta_{11}E_1 + \beta_{12}E_2)S_1 - \gamma_1E_1 + f_1'(1 - \gamma_2)E_2 - (b_1' + \mu_1')(1 - \gamma_1)E_1 - r_1E_1 \\
&\quad + K_1I_1 - r_{12}E_1I_2 + f_1''K_2I_2 \qquad (3) \\
\frac{\partial E_2}{\partial t} &= (\beta_{21}E_1 + \beta_{22}E_2)S_2 - \gamma_2E_2 - (\mu_2' + b_2')(1 - \gamma_2)E_2 - r_2E_2 + b_1'(1 - \gamma_1)E_1 \\
&\quad + K_2I_2 - r_{23}E_2I_3 \\
\frac{\partial E_3}{\partial t} &= b_2'(1 - \gamma_2)E_2 + \beta_{33}S_3E_3 - (\mu_3' + \gamma_3 + r_3)E_3 + K_3I_3 + K_{23}I_2E_3 + \beta_{23}S_2E_3 \\
\frac{\partial I_1}{\partial t} &= r_1E_1 - K_1I_1 + f_1''(1 - K_2)I_2 - (b_1'' + \mu_1'')(1 - K_1)I_1 - K_{12}I_1E_2
\end{aligned}
$$

$$\frac{\partial I_2}{\partial t} = r_2E_2 + b_1''(1-K_1)I_1 - K_2I_2 - (b_2'' + \mu_2'')(1-K_2)I_2 - K_{23}I_2E_3$$

$$\frac{\partial I_3}{\partial t} = b_2''(1-K_2)I_2 + r_3E_3 + K_{23}E_2I_3 - (K_3 + \mu_3'')I_3$$

5 The Copula Approach

We propose now to introduce the statistical notion of copula in order to check rapidly the nature of the distribution of the sizes of the subpopulations involved in the epidemic process as well as the stochastic dependency between these sizes considered as random variables.

Proposition 1 *Assume that there exist three age classes into the host subpopulations whose sojourn times T_i for $i = 1, 2, 3$, are independent random variables defined on the probabilized space $(\Omega, \mathfrak{F}, \mathbb{P})$, then we can relate the survival functions S_j for $j = 1, 2, 3$, by:*

$$\mathbb{P}\left(T_i > t_i \ \textit{for}\ i = 1, 2, 3\right) = \exp\left[-\left(\sum_{j=1,2,3}\left(-\ln\left(S_j(t_j)\right)^{\frac{1}{\alpha}}\right)\right)^{\alpha}\right]$$

$$\text{for} \quad t_1, t_2, t_3 > 0$$

where α is a parameter of dependence.

Proof We define the mean survival function, expectation of the survival function $S(t, q)$, i.e., the probability to survive until the age t within a random risk q, by:

$$S(t) = \mathbb{E}_q\left[B(t)^q\right]$$

where B is a decreasing function (e.g., $B(t) = \exp(-t)$ if T is an exponential random variable) and $\mathbb{E}_q$ denotes the conditional expectation relatively to q.

Recall that the Laplace transform of a positive random variable q is defined by:

$$L_q(s) = \mathbb{E}_q\left[\exp\left(-sq\right)\right] = \int_{R+} \exp\left(-st\right) dG_q(t)$$

where G_q is the distribution function of q. It is also the generating function evaluated at $\ln(-s)$; thus, the knowledge of $L_q(s)$ determines entirely the distribution of q.

Using the Laplace transformation, we obtain:

$$\mathbb{E}_q\left(\exp(-sq)\right) = \exp\left(-s^{\alpha}\right) \tag{4}$$

On the other hand, we have that:

$$\begin{aligned}
\mathbb{P}\left(T_i > t_i \text{ for } i = 1,2,3\right) &= \mathbb{E}_q\left[\prod_{j=1,2,3} B_j(t_j)^q\right] \\
&= \mathbb{E}_q\left[\prod_{j=1,2,3} \exp\left(q \ln\left(B_j(t_j)\right)\right)\right] \\
&= \mathbb{E}_q\left[\exp\left(\sum_{j=1,2,3} q \,\ln\left(B_j(t_j)\right)\right)\right]
\end{aligned}$$

From the Eq. (4), we have:

$$\mathbb{P}\left(T_i > t_i \;\text{ for } i = 1,2,3\right) = \exp\left[-\left(-\sum_{j=1,2,3} \ln\left(B_j(t_j)\right)\right)^{\alpha}\right]$$

We have indeed:

$$S_i(t_i) = \exp\left[-\left(-\ln\left(B_j(t_j)\right)\right)^{\alpha}\right]$$

Thus:

$$\left(-\sum_{j=1,2,3} \ln\left(B_j(t_j)\right)\right)^{\alpha} = \sum_{j=1,2,3}\left(-\ln\left(S_j(t_j)\right)^{\frac{1}{\alpha}}\right)^{\alpha}$$

Therefore:

$$\mathbb{P}\left(T_i > t_i \text{ for } i = 1,2,3\right) = \exp\left[-\left(\sum_{j=1,2,3}\left(-\ln\left(S_j(t_j)\right)^{\frac{1}{\alpha}}\right)\right)^{\alpha}\right]$$

■

On another side, we defined the Archimedean copula as follows:

$$C(u_1,\ldots,u_n) = \begin{cases} \phi^{-1}\left(\phi(u_1) + \cdots + \phi(u_n)\right) & \text{if} \quad \phi(u_1) + \cdots + \phi(u_n) \le 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where the generator of the copula ϕ is a twice continuously differentiable function which satisfies:

$$\phi(1) = 0, \;\; \phi^{(1)}(u) < 0 \quad \textit{and} \quad \phi^{(2)}(u) > 0 \quad \textit{for all} \quad u \in [0,1]^n$$

where $\phi^{(i)}$ denotes the i-th order derivative of ϕ.

Notice that a popular Archimedean copula is the Gumbel–Hougaard copula that is defined as follows:

$$\mathbf{C}(u_1, \ldots, u_n) = \exp\left\{-\left[(-\ln(u_1))^\alpha + \cdots + (-\ln(u_n))^\alpha\right]^{\frac{1}{\alpha}}\right\} \tag{6}$$

where $\alpha \geq 1$ and $\phi(t) = (-\ln t)^\alpha$.

So from Proposition 1 and the Eq. (6), we obtain:

$$\begin{aligned}\mathbb{P}\left(T_i > t_i \text{ for } i = 1,2,3\right) &= \exp\left[-\left(\sum_{j=1,2,3}\left(-\ln\left(S_j(t_j)\right)^{\frac{1}{\alpha}}\right)\right)^\alpha\right] \\ &= \mathbf{C}(S_1, S_2, S_3) \end{aligned} \tag{7}$$

where $\mathbf{C}$ is a Gumbel–Hougaard copula.

6 Results

The theoretical study of the dynamics of the differential systems presented in the previous sections can be found in [8–10]. We will focus here only on simulation results. We have used the classical Gillespie IBM (Individual-Based Modelling) approach for simulating the stochastic version [8–11] of the above differential systems and we focus on the use of the copula approach in the study of the stochastic dependency between various sizes of subpopulations presented above, considered as random variables.

The Figs. 4 and 5 represent respectively the evolution of the sizes of different compartments of the models (2) and (3) and the joined distribution of some couples of these variables. We can notice that as expected the young subjects infected have less influence on the adults susceptible than the adults infected on the young subjects susceptible. The parameters used for obtaining this result are:

$$f_1 = 75k, f_2 = 25k, \delta_1 = 2k/3, \delta_2 = 4k/5, b = 98k/96, l_1 = 30k, l_2 = 101k,$$

$$v = 70k/96, e_1 = k, e_2 = 6k/5, h_1 = 15k, h_2 = 3k, u = 50k/96, c_1 = 4k/3, c_2 = 2k,$$

$$\beta_1 = 4k/100, \beta_2 = k/10, K = 9k/10, r = 2k/5, k = 1.$$

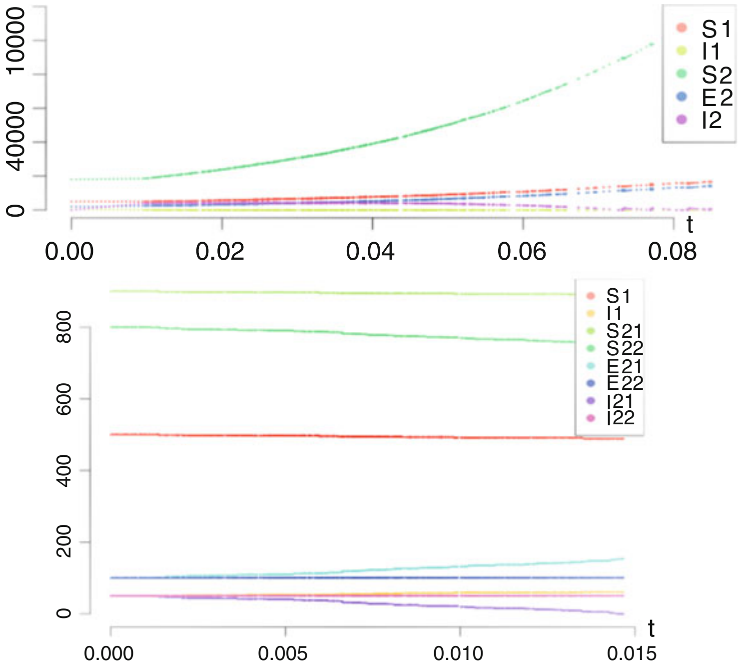

Fig. 4 Simulation of the variables of the stochastic version of the model (3) in case of demographic increase (*top*) and of the model (2) in case of demographic decrease (*bottom*), using the Gillespie R-package http://www.inside-r.org/node/46167

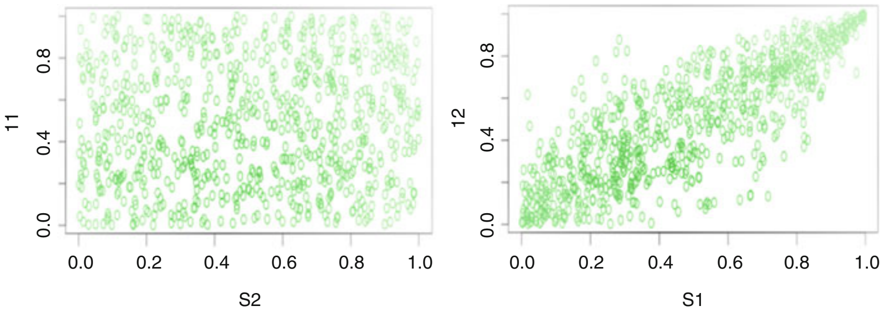

Fig. 5 Distribution function of the couple (S_2, I_1) (*left*) and of the couple (S_1, I_2) (*right*) in model (3). Taken from [11]. Published with the kind permission of © American Institute of Mathematical Sciences 2013. All rights reserved

If C denotes the Gumbel copula, we recall that the value of the p-quantile of the distribution of the expectation of E_2 conditionally to $S_2 = s$ is given by the following formula [11]:

$$p = C(f(s), gs(p))(\ln(f(s))/\ln[C(f(s), gs(p))])^{\alpha-1}/f(s)$$

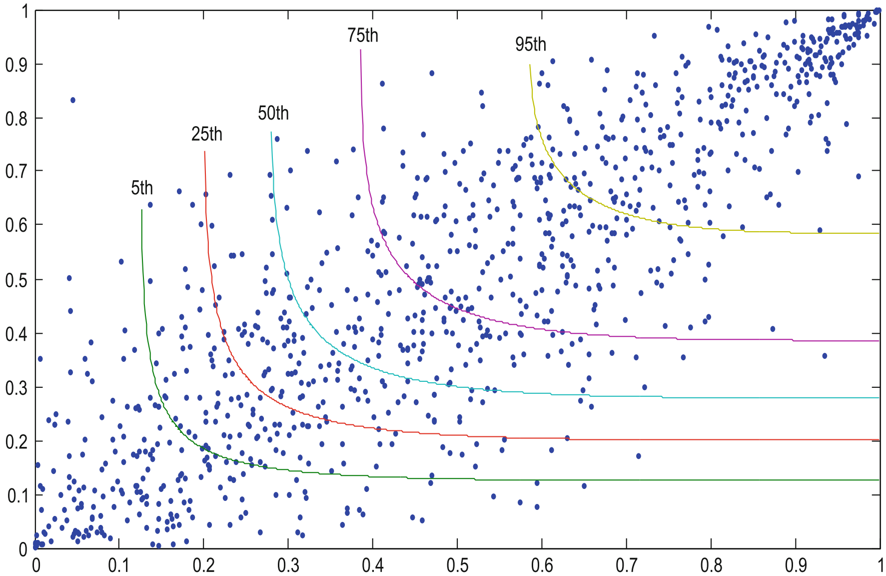

Fig. 6 Study of the interaction between the sizes S_2 and E_2 in model (3), by using the Gumbel copula with $\alpha = 3$ for representing the p-quantile regression curves of E_2 conditionally to S_2, for different values of p (from 5 to 95), superimposed on the scatter plot of the joined distribution of the couple (S_2, E_2). Taken from [11]. Published with the kind permission of

where f and gs are respectively the distribution functions of S_2 and of E_2 conditionally to $S_2 = s$.

We have superimposed in Fig. 6 the quantile regression curves for different values of p on the scatter plot of the two distribution functions S_2 and E_2. These curves permit to divide the population into several parts where the individuals in each part are dependent. Then we can analyze each part to make explicit dependencies between all individuals in the population. It is important for the justification of the derivation of the Ross–McKendrick modelling from an underlying stochastic process simulated by IBM approach, because it asks in principle the absence of correlation between susceptible and infected in order to get the contagious quadratic term in the differential equations [8–11]. We see on the given example that this hypothesis is roughly available only for the week values of the conditional quantile p.

7 Conclusion

We have considered some natural extensions of the classical Ross–McKendrick–Macdonald approaches, adding the age classes of the human host. Two examples have been presented, which show the interest of the introduction of age classes into the classical equation, by presenting the interaction graph for each model. In the future, we will study several demographic and risk models like the Usher model and the Cox model, in order to perform the copula approach with the various presented epidemic model. This copula approach allows us to find the relationships between the different classes, in order to see how we can reduce the infectious contacts.

Acknowledgements We indebted to Campus France CMCU for supporting us with the grant PHC Maghreb SCIM.

References

1. Demongeot, J.: Biological boundaries and biological age. Acta Biotheor. **57**, 397–419 (2009)
2. Gaudart, J., Ghassani, M., Mintsa, J., Waku, J., Rachdi, M., Doumbo, O.K., Demongeot, J.: Demographic and spatial factors as causes of an epidemic spread, the copula approach. In: IEEE Advanced Information Networking and Application (AINA'10, Perth, April 2010). IEEE Proceedings, Piscataway, pp. 751–758 (2010)
3. Demongeot, J., Waku, J.: Counter-examples for the population size growth in demography. Math. Popul. Stud. **12**, 199–210 (2005)
4. Cox, W.O.: Regression models and life-tables. J. R. Stat. Soc. Ser. B Methodol. **34**, 187–220 (1972)
5. Macdonald, G.: The Epidemiology and Control of Malaria. Oxford University Press, London (1957)
6. Nelsen, R.B.: Copulas and association. In : Advances in Probability Distributions with Given Marginals. Beyond the Copulas, pp. 51–74. Kluwer, Dordrecht (1991)
7. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics II. The problem of endemicity. Proc. R. Soc. Lond. Ser. A **138**, 55–83 (1932)
8. Gaudart, J., Ghassani, M., Mintsa, J., Rachdi, M., Waku, J., Demongeot, J.: Demography and diffusion in epidemics: Malaria and black death spread. Acta Biotheor. **58**, 277–305 (2010)
9. Demongeot, J., Gaudart, J., Mintsa, J., Rachdi, M.: Demography in epidemics modelling. Commun. Pure Appl. Anal. **11**, 61–82 (2012)
10. Demongeot, J., Gaudart, J., Lontos, A., Promayon, E., Mintsa, J., Rachdi, M.: Least diffusion zones in morphogenesis and epidemiology, Int. J. Bifurcation Chaos **22**, 1250028 (2012)
11. Demongeot, J., Ghassani, M., Rachdi, M., Ouassou, I., Taramasco, C.: Archimedean copula and contagion modeling in epidemiology. Netw. Heterogeneous Media **8**, 149–170 (2013)

Index

E. Ould Saïd et al. (eds.), *Functional Statistics and Applications*,
Contributions to Statistics, DOI 10.1007/978-3-319-22476-3