

Nonparametric estimation of a quantile density function by wavelet methods

Christophe Chesneau, Isha Dewan, Hassan Doosti

▶ **To cite this version:**

Christophe Chesneau, Isha Dewan, Hassan Doosti. Nonparametric estimation of a quantile density function by wavelet methods. 2015. hal-01153638v3

HAL Id: hal-01153638
https://hal.science/hal-01153638v3

Preprint submitted on 10 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nonparametric estimation of a quantile density function by wavelet methods

Christophe Chesneau*

Laboratoire de Mathématiques Nicolas Oresme, Université de Caen, FRANCE.

Isha Dewan

Indian Statistical Institute, New Delhi, INDIA.

Hassan Doosti

Mashhad University of Medical Sciences, Mashhad, IRAN.

Abstract

In this paper nonparametric wavelet estimators of the quantile density function are proposed. Consistency of the wavelet estimators is established under the $\mathbb{L}_p$ risk. A simulation study illustrates the good performance of our estimators.

Keywords: Quantile Density Estimation, Rates of convergence, Wavelet methods

2010 MSC: 62G07, 62G20

1. Introduction

Motivation. Quantile density function, the derivative of the quantile function, comes up in the study of lifetime and survival data. The expression for the limiting variance of empirical and kernel type estimators of the quantile function involves the quantile density function. The hazard quantile function too can be written in terms of the quantile density function. Nonparametric estimators of the quantile density function have been studied by [33], [23] and [45]. Most of these estimators underperform at the tails, see, for example, [45]. Hence there is a need to look at alternate estimators.

On the quantile function. Quantiles are often used for statistical modeling and analysis

Email addresses: christophe.chesneau@unicaen.fr (Christophe Chesneau*), ishadewan@gmail.com (Isha Dewan), doostih@mums.ac.ir (Hassan Doosti)

of data. Measures based on quantiles are less influenced by outliers. Hence they are particularly useful in the study of lifetime data and also for studying heavy tailed distributions. Sometimes the distribution function of the random variable of interest can not be expressed in a closed form. However, the quantile function can be written in an explicit form, for example, Generalized Lambda distribution (GLD) and Skew logistic distribution (see [20] and [24]).

Researchers have used quantiles for preliminary data analysis. Statistical analysis of data based on quantile functions has been carried out in reliability and survival analysis and other branches of applied statistics (see, for example, [37], [44], [48], [30], [31], [39] and [41]). [46, 47] have developed multiple comparison procedures for quantile functions. Nonparametric test procedures under competing risks have been developed by [34], [22] and [40].

Let X be a continuous random variable with cumulative distribution function $F(x)$, density function $f(x)$ and hazard function $r(x)$. The quantile function of X is defined as

$$Q(x) = F^{-1}(x) = \inf\{y \in R;\ F(y) \geq x\}, \qquad x \in [0,1]. \tag{1.1}$$

It satisfies

$$F(Q(x)) = x. \tag{1.2}$$

[32] and [23] defined the quantile density function corresponding to quantile function $Q(x)$ by

$$g(x) = Q'(x) = \frac{1}{f(Q(x))}, \qquad x \in [0,1]. \tag{1.3}$$

Note that the sum of two quantile density functions is again a quantile density function. This idea is useful in modeling data. [31] defined the hazard quantile function $R(x)$ as follows :

$$R(x) = r(Q(x)) = \frac{f(Q(x))}{1 - F(Q(x))} = \frac{1}{(1-x)g(x)}, \qquad x \in (0,1).$$

Hence a nonparametric estimator of the quantile density function will give us an estimator of the hazard quantile function.

Overview of previous works. Let $X_1, X_2, \ldots, X_n$ be *iid* random variables from distribution $F(x)$ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. [23] suggested the following two smooth estimators of the quantile density function. The first one, denoted by $\hat{g}_{j1}(x)$, is given by

$$\hat{g}_{j1}(x) = \frac{1}{\hat{f}(\hat{Q}(x))}, \qquad x \in [0,1], \tag{1.4}$$

where $\hat{f}(x)$ is a kernel type density estimator of the form :

$$\hat{f}(x) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\Big(\frac{x - X_i}{h(n)}\Big),$$

where $h(n)$ is the bandwidth and $K(.)$ an appropriate kernel function, and $\hat{Q}(x)$ is the usual empirical estimator of $Q(x)$. The properties that the kernel function and the bandwidth need to satisfy have been listed in [45].

The second estimator of [23] is given as

$$\hat{g}_{j2}(x) = \sum_{i=2}^{n} X_{(i)} \Big(K_{h(n)}\Big(x - \frac{i-1}{n}\Big) - K_{h(n)}\Big(x - \frac{i}{n}\Big)\Big), \qquad x \in [0,1], \tag{1.5}$$

where $X_{(i)}$ is the i^{th}order statistic, $i = 1, 2, \ldots, n$, and $K_h(u) = (1/h)K(u/h)$.

[23] and [45] showed that the performance of the estimator of $g(x)$ given by (1.4) is better than that of the estimator given by (1.5).

[45] proposed the following smooth estimator of the quantile density function :

$$\hat{g}_S(x) = \frac{1}{h(n)} \int_0^1 K\Big(\frac{t-x}{h(n)}\Big) \frac{1}{\hat{f}(\hat{Q}(t))} dt, \qquad x \in [0,1], \tag{1.6}$$

where $K(.)$ is a kernel and $h(n)$ is the bandwidth sequence. It can be also be expressed as

$$\hat{g}_S(x) = \frac{1}{nh(n)} \sum_{i=1}^{n} K\Big(\frac{S_i - x}{h(n)}\Big) \frac{1}{\hat{f}(X_{(i)})}.$$

where S_i is the proportion of observations less than or equal to $X_{(i)}$, the i^{th} order statistic.

Contributions. In this paper, we explore a different approach by considering estimators based on projections on a wavelet basis of the (crude) form :

$$\hat{g}(x) = \sum_{m} \hat{a}_m e_m(x), \qquad x \in [0,1],$$

where $\hat{a}_m$ denotes an estimator of the coefficient $a_m = \int g(x) e_m(x) dx$ and $\{e_m\}$ forms the wavelet basis. Such basis is of interest, thanks to its localization in space and frequency properties. For the standard nonparametric estimation problems in density, regression,.... wavelet estimators outperform kernel estimators in representing discontinuities (edges, spikes,...). Basics on wavelet estimation can be found in [21].

In this study, we develop two kinds of wavelet estimators for the quantile density function g : a linear one based on simple projections and a nonlinear one based on a hard thresholding rule. Our

wavelet hard thresholding estimator has the feature to be adaptive according to $g(x)$. Let us mention that, due to the choice of $\hat{a}_m$ considered, our estimators belong to the family of warped wavelet basis estimators introduced by [26] in another statistical context - regression problem with random design. Under some smoothness assumptions on $g(x)$, we determine fast rates of convergence of the proposed estimators under the $\mathbb{L}_p$ risk.

Paper organization. The rest of this paper is organized as follows. In the next section, we present our wavelet estimators. The main theoretical results are described in Section 3 and Section 4 is devoted to the numerical performances of our estimators. The proofs of the technical results appear in Sections 5.

2. Wavelet estimators

Wavelet expansion. We define the spaces $\mathbb{L}_p([0,1])$, $p \geq 1$, by

$$\mathbb{L}_p([0,1]) = \left\{ h : [0,1] \to R;\ \int_{[0,1]} |h(x)|^p dx < \infty \right\}.$$

For the purpose of the paper, we consider an orthonormal wavelet basis of $\mathbb{L}_2([0,1])$ associated with the initial wavelet functions ϕ and ψ of the Daubechies wavelets $db2N$, where N denotes a positive integer. The functions ϕ and ψ are compactly supported and continuously differentiable.

For any $x \in R$,

$$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k).$$

Then, with an appropriate treatment at the boundaries, there exists an integer τ satisfying $2^\tau \geq 2N$ such that, for any $\ell \geq \tau$, the system

$$\mathcal{W} = \{\phi_{\ell,k};\ k \in \{0,\dots,2^\ell - 1\};\ \psi_{j,k};\ k \in \{0,\dots,2^j - 1\},\ j \geq \ell\}$$

is an orthonormal basis of $\mathbb{L}_2([0,1])$.

For convenience, we suppose that X_1 is compactly supported, say $X_1(\Omega) = [0,1]$, and that F is continuous and strictly monotone from $[0,1]$ to $[0,1]$.

Let us suppose that $g \in \mathbb{L}_2([0,1])$. Hence, for any integer $\ell \geq \tau$, we have the following wavelet expansion on $\mathcal{W}$:

$$g(x) = \sum_{k=0}^{2^\ell - 1} c_{\ell,k}\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x), \qquad x \in [0,1],$$

where

$$c_{j,k} = \int_{[0,1]} g(x)\phi_{j,k}(x)dx, \qquad d_{j,k} = \int_{[0,1]} g(x)\psi_{j,k}(x)dx. \tag{2.1}$$

All the technical details can be found in, e.g., [11] and [27].

Wavelet coefficients estimators. The wavelet coefficients $c_{j,k}$ and $d_{j,k}$ are unknown and need to be estimated. Our approach is based on the following remark : by the change of variable $x = F(y)$ with $y \in X_1(\Omega) = [0,1]$, we can rewrite $c_{j,k}$ as

$$\begin{aligned} c_{j,k} &= \int_{[0,1]} g(x)\phi_{j,k}(x)dx = \int_{[0,1]} \frac{1}{f(Q(x))}\phi_{j,k}(x)dx \\ &= \int_{[0,1]} \phi_{j,k}(F(x))dx. \end{aligned} \tag{2.2}$$

Similarly

$$d_{j,k} = \int_{[0,1]} \psi_{j,k}(F(x))dx. \tag{2.3}$$

Since F is unknown, we estimate it by the empirical estimator :

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \leq x\}}, \qquad x \in [0,1],$$

where $\mathbf{1}$ is the indicator function.

This leads to the following integral estimator for $c_{j,k}$:

$$\hat{c}_{j,k} = \int_{[0,1]} \phi_{j,k}(\hat{F}(x))dx. \tag{2.4}$$

The analogous estimator for $d_{j,k}$ is

$$\hat{d}_{j,k} = \int_{[0,1]} \psi_{j,k}(\hat{F}(x))dx. \tag{2.5}$$

Due to the composition of the element of the wavelet basis with $\hat{F}(x)$, $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ can be viewed as warped wavelet basis coefficient estimators. Such estimators were introduced by [26] in the context of regression with random design. Other improvements and modern developments can be found in [5, 6] for the same regression model and in [7] for the relative density estimation problem (with trigonometric basis).

Remark 2.1. *Clearly, $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ are not unbiased estimators for $c_{j,k}$ and $d_{j,k}$. However, using the dominated convergence theorem, one can prove that they are asymptotically unbiased. Moreover, they satisfy power moments and concentration inequalities. See Lemmas 5.1, 5.2, 5.3 and 5.4 below.*

Remark 2.2. *Our study can be extended for $X_1(\Omega) = [a, b]$ with $a < b$ finite. Using the change of variables $x = F(y)$ with $y \in X_1(\Omega) = [a, b]$, we get*

$$c_{j,k} = \int_{[0,1]} g(x)\phi_{j,k}(x)dx = \int_{[a,b]} \phi_{j,k}(F(x))dx.$$

So we consider the last integral over $[a, b]$ instead of $[0, 1]$.

Based on $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ given in (2.4) and (2.5), respectively, we consider two kinds of wavelet estimators for $g(x)$: a linear wavelet estimator $\hat{g}_L(x)$ and a hard thresholding wavelet estimator $\hat{g}_H(x)$, both defined below.

Linear wavelet estimator. We define the linear wavelet estimator $\hat{g}_L(x)$ by

$$\hat{g}_L(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}\phi_{j_0,k}(x), \qquad x \in [0, 1], \tag{2.6}$$

where j_0 is a positive integer chosen a posteriori (see Theorems 3.1 and 3.3).

Hard thresholding wavelet estimator. We define the hard thresholding wavelet estimator $\hat{g}_H(x)$ by

$$\hat{g}_H(x) = \sum_{k=0}^{2^{\tau}-1} \hat{c}_{\tau,k}\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \hat{d}_{j,k}\mathbf{1}_{\{|\hat{d}_{j,k}|\geq\lambda_j\}}\psi_{j,k}(x), \qquad x \in [0, 1], \tag{2.7}$$

where $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ are defined by (2.4) and (2.5), j_1 is a positive integer and λ_j represents a threshold. Both j_1 and λ_j will be chosen a posteriori (see Theorems 3.2 and 3.4)

The construction of $\hat{g}_H(x)$ exploits the sparse nature of the wavelet decomposition of $g(x)$: only the wavelet coefficients $d_{j,k}$ with large magnitude contain the main information (in terms of details) of g. Hence $\hat{g}_H(x)$ aims to estimate only the larger coefficients and remove the other (or estimate it by 0). Further aspects and explanation related to this selection techniques can be found in [3], [21] and [50].

As usual in wavelet estimation, we will suppose that the unknown function $g(x)$ belongs to Besov balls defined below.

Besov balls. Let $M > 0$, $s > 0$, $r \geq 1$ and $q \geq 1$. We say that $g(x)$ belongs to the Besov ball $B^s_{r,q}(M)$ if there exists a constant $M^* > 0$ (depending on M) such that (2.1) satisfy

$$\left(\sum_{k=0}^{2^\tau - 1} |c_{\tau,k}|^r \right)^{1/r} + \left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/r)} \left(\sum_{k=0}^{2^j-1} |d_{j,k}|^r \right)^{1/r} \right)^q \right)^{1/q} \leq M^*,$$

with the usual modifications if $r = \infty$ or $q = \infty$.

In this expression, s is a smoothness parameter and r and q are norm parameters. Details on Besov balls can be found in [29] and [21, Chapter 9].

3. Theoretical results

3.1. First results

This section is devoted to the study of performance of wavelet estimators $\hat{g}_L(x)$ and $\hat{g}_H(x)$.

Theorem 3.1 determines the rates of convergence attained by $\hat{g}_L(x)$ under the $\mathbb{L}_p$-risk over Besov balls.

Theorem 3.1. *Let $p \geq 1$. Assume that $g \in \mathbb{L}_{\max(r,2)}([0,1]) \cap B^s_{r,q}(M)$ with $s > 1/r$, $r \geq 1$ and $q \geq 1$. Set $s_* = \min(s, s - 1/r + 1/p)$ and let $\hat{g}_L(x)$ be as in (2.6) with j_0 being the integer such that*

$$n^{1/(2s_*+4)} < 2^{j_0+1} \leq 2n^{1/(2s_*+4)}.$$

Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_L(x) - g(x)|^p dx \right) \leq C n^{-s_* p/(2s_*+4)}.$$

The proof is based on statistical properties of $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ (see Lemmas 5.1 and 5.2), and technical bounds related to wavelet series and the $\mathbb{L}_p$ norm. At this stage, let us consider the rate of convergence $n^{-s_* p/(2s_*+4)}$ as a benchmark. This aspect will be discussed later.

Theorem 3.2 explores the rates of convergence of $\hat{g}_H(x)$ under the $\mathbb{L}_p$-risk over Besov balls.

Theorem 3.2. *Let $p \geq 1$ and $\hat{g}_H$ be as in (2.7) with j_1 being the integer satisfying*

$$\left(\frac{n}{\ln(n)}\right)^{1/4} < 2^{j_1+1} \leq 2\left(\frac{n}{\ln(n)}\right)^{1/4},$$

and λ_j being the threshold :

$$\lambda_j = K 2^{3j/2}\sqrt{2p}\sqrt{\frac{\ln(n)}{n}}, \tag{3.1}$$

with $K = \sup_{x\in[0,1]} |\psi'(x)|$.

Suppose that $g \in \mathbb{L}_{\max(r,2)}([0,1]) \cap B^s_{r,q}(M)$ with $s > 1/r$, $r \geq 1$ and $q \geq 1$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_H(x) - g(x)|^p dx\right) \leq C\varphi_n,$$

where

$$\varphi_n = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{sp/(2s+4)}, & \text{for} \quad rs > 2(p-r), \\ \left(\dfrac{\ln(n)}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+4)}, & \text{for} \quad rs < 2(p-r), \\ \left(\dfrac{\ln(n)}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+4)} (\ln(n))^{(p-r/q)_+}, & \text{for} \quad rs = 2(p-r). \end{cases}$$

The proof is based on statistical properties of $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ (see Lemmas 5.3 and 5.4), and a general result on the $\mathbb{L}_p$ risk of the hard thresholding wavelet estimator which can be proved by combining Theorem 5.1 of [25] and Theorem 4.2 of [8].

If we do a global comparison between the results of Theorems 3.1 and 3.2, the rates of convergence achieved by $\hat{g}_H(x)$ are better than the one achieved by $\hat{g}_L(x)$. Moreover, let us recall that $\hat{g}_H(x)$ is adaptive while $\hat{g}_L(x)$ is not adaptive due to its dependence on s in its construction.

In comparison to the standard density estimation problem, the rates of convergence obtained in Theorems 3.1 and 3.2 are slower. To be more specific,

- for the wavelet linear estimation, the standard rate of convergence is $n^{-s_*p/(2s_*+1)}$, against $n^{-s_*p/(2s_*+4)}$ in Theorem 3.1 (see [21, Section 10.2]).
- for the wavelet nonlinear estimation, the standard rate of convergence is similar to φ_n but with "+1" instead of "+4" in the exponent (see [13]).

The rest of the study proves that, under an additional assumption on $g(x)$, the rates of convergence obtained above can be improved and be made equal to the standard one.

3.2. Improved results but with an additional assumption

We now introduce the following Lipschitz(1/2) assumption :

(A) there exists a constant $C_* > 0$ such that, for any $(x, y) \in [0, 1]^2$,

$$|g(x) - g(y)| \leq C_* \sqrt{|x - y|}.$$

Theorems 3.3 and 3.4 below show that, under **(A)**, one can construct more efficients linear and hard thresholding wavelet estimators than those presented in Theorems 3.1 and 3.2.

Theorem 3.3. *Let $p \geq 1$. Suppose that* **(A)** *is satisfied and $g \in B^s_{r,q}(M)$ with $s > 1/r$, $r \geq 1$ and $q \geq 1$ such that $\min(s, s - 1/r + 1/p) > 1/2$. Set $s_* = \min(s, s - 1/r + 1/p)$ and $\hat{g}_L(x)$ be as in (2.6) with j_0 being the integer such that*

$$n^{1/(2s_*+1)} < 2^{j_0+1} \leq 2n^{1/(2s_*+1)}.$$

Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_L(x) - g(x)|^p dx\right) \leq C n^{-s_* p/(2s_*+1)}.$$

The proof is similar to the one of Theorem 3.1. The techniques involved in the proof make use of statistical results on $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ derived by [26] for the estimation of the regression function . These statistical results justify the consideration of **(A)** and a restriction on j_0, i.e., $2^{j_0} \leq \sqrt{n/\ln(n)}$.

Theorem 3.4. *Let $p \geq 1$. Suppose that* **(A)** *is satisfied. Let $\hat{g}_H(x)$ be (2.7) with j_1 being the integer satisfying*

$$\sqrt{\frac{n}{\ln(n)}} < 2^{j_1+1} \leq 2\sqrt{\frac{n}{\ln(n)}},$$

and λ_j being the threshold :

$$\lambda_j = \kappa \sqrt{\frac{\ln(n)}{n}}, \tag{3.2}$$

where κ is a large enough constant (depending, among other, on C_* and ψ). Suppose that $g \in B^s_{r,q}(M)$ with $s > 1/r$, $r \geq 1$ and $q \geq 1$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_H(x) - g(x)|^p dx\right) \leq C\Psi_n,$$

where

$$\Psi_n = \begin{cases} \left(\dfrac{\ln(n)}{n}\right)^{sp/(2s+1)}, & \textit{for} \quad 2rs > (p-r), \\ \left(\dfrac{\ln(n)}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+1)}, & \textit{for} \quad (r,s,p) \in \mathcal{D}, \end{cases}$$

where $\mathcal{D} = \{(r,s,q) \in [1,\infty) \times (1/r,\infty) \times [1,\infty);\ 1/2 + 1/r \leq s \leq (p-r)/(2r),\ r \leq (p-2)/(2(s - 1/r) + 1)\}$.

The proof uses arguments similar to those in Theorem 3.2. Again, the main originality is the use of new statistical results on $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ derived from the results proved by [26].

Let us remark that Ψ_n corresponds to the standard rate of convergence for the standard density estimation problem via the hard thresholding wavelet method (see [13]).

The next section explores the numerical performances of our wavelet estimators and compares them with the estimators developed by [23] and [45].

4. Simulation Study

We consider the Beta distribution and the Generalized Lambda Distributions (GLD). It is easy to see that when parameters in Beta distribution were chosen from $(0,1)$, the corresponding quantile density function satisfies all the conditions required to prove the results. On the other hand GLD has four different parameters, *i.e.*, $\lambda_1, \lambda_2, \lambda_3$ and λ_4. The parameters λ_1 and λ_2 are, respectively, location and scale parameters, while λ_3 and λ_4 determine the skewness and kurtosis of the GLD$(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, see [24]. [19, Section 3] noted that GLD is very rich in the variety of density and tail shapes. It contains unimodal, U-shaped, J-shaped and monotone probability density functions. These can be symmetric or asymmetric and their tails can be smooth, abrupt, or truncated, and long, medium or short depending on the choice of parameters. [24, page 43] showed that when $\lambda_3 > 0$ and $\lambda_4 > 0$ the support of GLD is $[\lambda_1 - 1/\lambda_2, \lambda_1 + 1/\lambda_2]$. On the other hand, it is easy to see from the closed form of the quantile function of GLD that when $\lambda_3 > 1$ and $\lambda_4 > 1$, we have $g \in \mathbb{L}_2([0,1])$.

We consider performance of linear wavelet estimator $\hat{g}_L(x)$ and the hard thresholding wavelet estimator $\hat{g}_H(x)$ presented in Section 2 and compare them with a linear wavelet estimator after local linear smoothing $\hat{g}_{LS}(x)$ proposed by [36]. The new smooth linear wavelet estimator has been adapted from [17, 18]. [1, 2], [9, 10] and [42, 43] have proposed this version of wavelet estimators

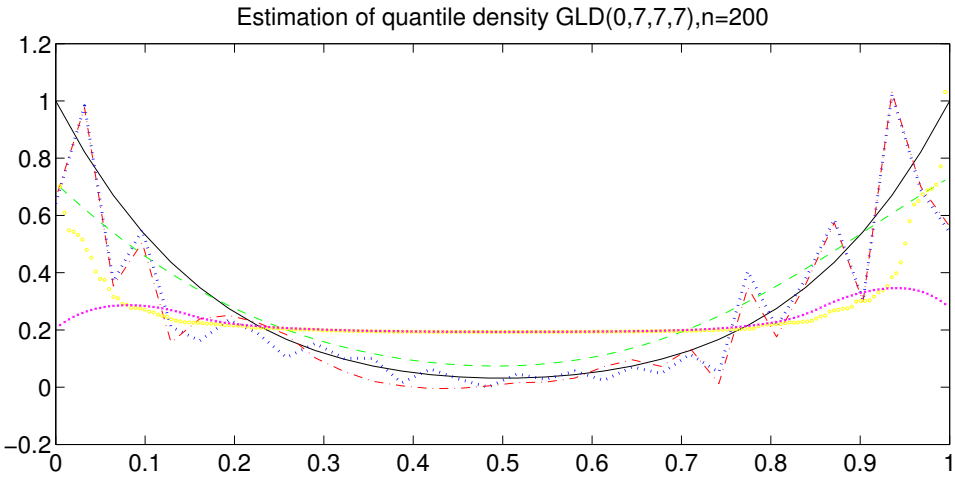


Figure 1: Estimation of quantile density. The black curve is the true GLD, the blue dotted line is linear wavelet estimator, $\hat{g}_L(x)$, the red dashed dot line is threshold wavelet estimator, $\hat{g}_H(x)$, the green dashed line is smooth version of our estimator, $\hat{g}_{LS}(x)$, the yellow line with circles is Jones' estimator, $\hat{g}_{j1}(x)$, and the magenta line with crosses is $\hat{g}_S(x)$.

and showed that it performs better than $\hat{g}_L(x)$ in different applications. Let us mention that it depends on a bandwidth $h(n)$ (equal to 0.15 if there is no mention). Integrals in wavelet coefficient estimators introduced in (2.4) and (2.5) have been approximated by Simpson's rule. Coarsest level, j_0, is chosen as 5. The parameters j_1 and λ_j are as given in Theorem 3.4 (according to the value of n). Following [36], the constant κ of threshold is chosen as a scaled absolute median of empirical wavelet coefficients. We employ Daubechies-Lagarias algorithm and write the codes in Matlab which are adapted from [36]. We compare these three estimators with $\hat{g}_{j1}(x)$ and $\hat{g}_S(x)$, two estimators given in (1.4) and (1.6), respectively. Following [45], the kernel function chosen is Triangular, $K(x) = (1 - |x|)\mathbf{1}_{\{|x|\leq 1\}}$ and $h(n) = 0.15$ if there is no mention.

Figures 1, 2 and 3 show the results of simulation for the Generalized Lambda Distributions : GLD(0,7,7,7), GLD(0.5,1,2,6) and the Beta distribution - Beta (0.5, 0.5), respectively. In each figure, the black curve is the true quantile density function, the blue dotted line is linear wavelet estimator $\hat{g}_L(x)$, the red dashed dot line is threshold wavelet estimator $\hat{g}_H(x)$, the green dashed line is smooth version of our estimator $\hat{g}_{LS}(x)$, the yellow line with circles is Jones' estimator $\hat{g}_{j1}(x)$ and the magenta line with crosses is Soni-Dewan-Jain's estimator $\hat{g}_S(x)$. From these three figures we conclude that the smooth version of the proposed wavelet estimators is closer to the unknown

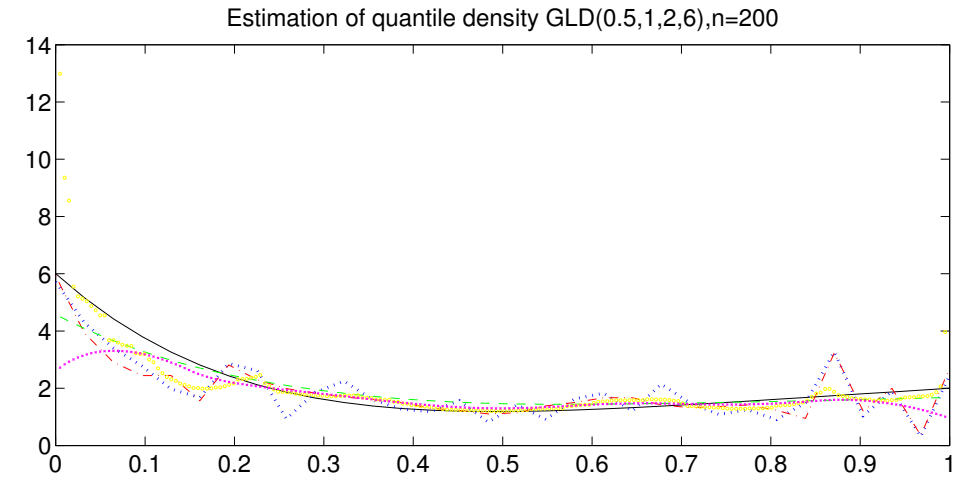


Figure 2: Estimation of quantile density. The black curve is the true GLD, the blue dotted line is linear wavelet estimator, $\hat{g}_L(x)$, the red dashed dot line is threshold wavelet estimator, $\hat{g}_H(x)$, the green dashed line is smooth version of our estimator, $\hat{g}_{LS}(x)$, the yellow line with circles is Jones' estimator, $\hat{g}_{j1}(x)$, and the magenta line with crosses is $\hat{g}_S(x)$.

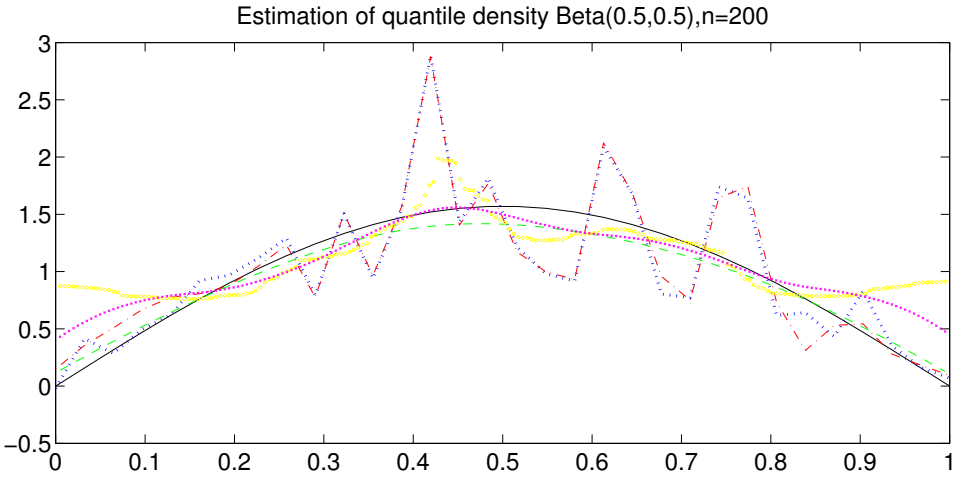


Figure 3: Estimation of quantile density. The black curve is the true Beta (0.5,0.5), the blue dotted line is linear wavelet estimator, $\hat{g}_L(x)$, the red dashed dot line is threshold wavelet estimator, $\hat{g}_H(x)$, the green dashed line is smooth version of our estimator, $\hat{g}_{LS}(x)$, the yellow line with circles is Jones' estimator, $\hat{g}_{j1}(x)$, and the magenta line with crosses is $\hat{g}_S(x)$.

Table 1: MISE for different estimators based on 500 replications

Estimator	GLD(0,7,7,7)		GLD(0.5,2,1.5,1.5)		GLD(0.1.5,1.5,1.5)		GLD(0.5,1,2,6)		Beta(0.5,0.5)	
	n=200	n=500	n=200	n=500	n=200	n=500	n=200	n=500	n=200	n=500
$\hat{g}_{LS}(x)$	**0.0078**	**0.0067**	**0.0101**	**0.0050**	**0.0189**	**0.0093**	**0.1149**	**0.0886**	**0.0252**	**0.0167**
$\hat{g}_{j1}(x)$	0.0268	0.0273	0.0648	0.0521	0.1088	0.0797	0.4718	0.2798	0.1259	0.1011
$\hat{g}_S(x)$	0.0512	0.0530	0.0127	0.0064	0.0256	0.0127	0.3758	0.3343	0.0604	0.0459

quantile density function as compared to the other four estimators we have studied.

Next we consider the mean integrated square error (MISE) of the smooth linear estimator of the quantile density function $\hat{g}_{LS}(x)$, $\hat{g}_{j1}(x)$ and $\hat{g}_S(x)$ for Beta distributions (Beta (0.5, 0.5)) and Generalized Lambda Distributions (GLD) GLD(0,7,7,7), GLD(0.5,2,1.5,1.5), GLD(0,1.5,1.5,1.5) and GLD(0.5,1,2,6).

Table 1 shows the estimation of Mean Integrated Square Error (MISE) for GLDs for four different choices of parameters and Beta(0.5, 0.5). The results in this Table 1 are based on 500 replications and sample sizes $n = 200$ and $n = 500$. We use the formula :

$$MISE = \frac{n-1}{500(n+1)^2} \sum_{i=1}^{500} \sum_{j=1}^{n} \left(\hat{f}_i \left(\frac{j}{n+1} \right) - f \left(\frac{j}{n+1} \right) \right)^2,$$

where $\hat{f}_i$ is the quantile density estimator in ith replication and f is true quantile density function.

Tables 2, 3 and 4 show the Mean Square Error (MSE) for different choices of fixed $x \in [0, 1]$ and different bandwidth parameters : $h(n) = 0.15, 0.19, 0.25$. We calculate the MSE of these estimators by generating 500 samples from GLD(0.5,1,2,6), Beta (0.5,0.5) and GLD(0,7,7,7), respectively. The sample size in each case is $n = 200$.

From Figures 1-3 and Tables 1-4 we conclude that

(i) our wavelet based estimator performs well in estimation of unknown quantile density. The performance of smooth version of linear wavelet estimator works the best.

(ii) our wavelet estimators perform well in the tails whereas [45] observed that the competitors

(1.4) and (1.6) suffer from boundary effect.

(iii) in majority of cases, the MSEs for the smooth version of the wavelet estimator are smaller than those of the competitors.

(iv) the MSEs and MISEs decrease as the sample size increases.

(v) the MSEs are bigger for x close to the tails.

(vi) the MISEs of smooth version of wavelet estimators are smaller than the MISEs of the competitors.

Table 2 MSEs (Standard Deviation) based on 500 replications and n=200, GLD(0.5,1,2,6)

		h(n)		
x	estimate	0.15	0.19	0.25
0.01	$\hat{g}_{LS}(x)$	1.5244(1.4414)	1.9204(1.4104)	2.5767(1.3743)
	$\hat{g}_{j1}(x)$	9.1864(21.656)	10.467(22.298)	10.115(20.877)
	$\hat{g}_{S}(x)$	9.4850(2.2242)	10.418(1.7664)	11.719(1.4807)
0.1045	$\hat{g}_{LS}(x)$	0.1027(0.1373)	0.1263(0.1635)	0.1595(0.1787)
	$\hat{g}_{j1}(x)$	0.9224(1.8540)	0.6914(1.2891)	0.4624(0.9149)
	$\hat{g}_{S}(x)$	0.1396(0.1615)	0.3028(0.2442)	0.7100(0.2998)
0.2040	$\hat{g}_{LS}(x)$	0.0609(0.0633)	0.0904(0.0727)	0.1278(0.0897)
	$\hat{g}_{j1}(x)$	0.2605(0.3657)	0.2045(0.3190)	0.1622(0.2456)
	$\hat{g}_{S}(x)$	0.0767(0.1019)	0.0536(0.0748)	0.0556(0.0595)
0.4030	$\hat{g}_{LS}(x)$	0.0682(0.0601)	0.1515(0.0758)	0.2929(0.0776)
	$\hat{g}_{j1}(x)$	0.0387(0.0638)	0.0322(0.0579)	0.0253(0.0460)
	$\hat{g}_{S}(x)$	0.0261(0.0405)	0.0313(0.0473)	0.0437(0.0513)
0.6020	$\hat{g}_{LS}(x)$	0.0165(0.0219)	0.0296(0.0263)	0.0580(0.0324)
	$\hat{g}_{j1}(x)$	0.0358(0.0548)	0.0293(0.0442)	0.0222(0.0423)
	$\hat{g}_{S}(x)$	0.0205(0.0270)	0.0173(0.0248)	0.0145(0.0234)
0.8010	$\hat{g}_{LS}(x)$	0.0127(0.0176)	0.0127(0.0176)	0.0134(0.0190)
	$\hat{g}_{j1}(x)$	0.0829(0.1376)	0.0566(0.0882)	0.0358(0.0540)
	$\hat{g}_{S}(x)$	0.0287(0.0361)	0.0183(0.0249)	0.0107(0.0147)
0.9005	$\hat{g}_{LS}(x)$	0.0478(0.0615)	0.0527(0.0672)	0.0553(0.0673)
	$\hat{g}_{j1}(x)$	0.0853(1403)	0.0600(0.0858)	0.0405(0.0953)
	$\hat{g}_{S}(x)$	0.0365(0.0499)	0.0431(0.0549)	0.0785(0.0634)
0.99	$\hat{g}_{LS}(x)$	0.1761(0.2068)	0.1566(0.1840)	0.1333(0.1587)
	$\hat{g}_{j1}(x)$	2.2551(2.7919)	2.2150(2.9676)	0.4694(2.3519)
	$\hat{g}_{S}(x)$	0.5055(0.2449)	0.5793(0.2201)	0.6505(0.1897)

Table 3 MSEs (Standard Deviation) based on 500 replications and n=200, Beta(0.5,0.5)

		h(n)		
x	estimate	0.15	0.19	0.25
0.01	$\hat{g}_{LS}(x)$	0.0477(0.0678)	0.0665(0.0816)	0.1253(0.1202)
	$\hat{g}_{j1}(x)$	0.7728(2429)	0.9684(0.2820)	1.2762(0.3435)
	$\hat{g}_{S}(x)$	0.1847(0.0532)	0.2204(0.0577)	0.2780(0.0649)
0.1045	$\hat{g}_{LS}(x)$	0.0231(0.0357)	0.0255(0.0399)	0.0414(0.0556)
	$\hat{g}_{j1}(x)$	0.0937(0.0580)	0.1663(0.0792)	0.3074(0.1169)
	$\hat{g}_{S}(x)$	0.0860(0.0619)	0.1083(0.0660)	0.1403(0.0700)
0.2040	$\hat{g}_{LS}(x)$	0.0142(0.0190)	0.0128(0.0174)	0.0125(0.0174)
	$\hat{g}_{j1}(x)$	0.0332(0.0409)	0.0175(0.0278)	0.0114(0.0197)
	$\hat{g}_{S}(x)$	0.0193(0.0252)	0.0151(0.0247)	0.0253(0.0347)
0.4030	$\hat{g}_{LS}(x)$	0.0288(0.0291)	0.0516(0.0309)	0.0912(0.0287)
	$\hat{g}_{j1}(x)$	0.0738(0.0412)	0.0471(0.0591)	0.0421(0.0565)
	$\hat{g}_{S}(x)$	0.0350(0.0446)	0.0342(0.0396)	0.0416(0.0403)
0.6020	$\hat{g}_{LS}(x)$	0.0365(0.0352)	0.0571(0.0338)	0.0999(0.0364)
	$\hat{g}_{j1}(x)$	0.0646(0.0888)	0.0502(0.0616)	0.0478(0.0663)
	$\hat{g}_{S}(x)$	0.0364(0.0457)	0.0317(0.0379)	0.0385(0.0405)
0.8010	$\hat{g}_{LS}(x)$	0.0169(0.0218)	0.0157(0.0212)	0.0151(0.0207)
	$\hat{g}_{j1}(x)$	0.0319(0.0507)	0.0132(0.0175)	0.0113(0.0178)
	$\hat{g}_{S}(x)$	0.0195(0.0297)	0.0176(0.0255)	0.0289(0.0333)
0.9005	$\hat{g}_{LS}(x)$	0.0136(0.0240)	0.0170(0.0250)	0.0210(0.0277)
	$\hat{g}_{j1}(x)$	0.1110(0.0618)	0.1994(0.0862)	0.3351(0.1155)
	$\hat{g}_{S}(x)$	0.0989(0.0658)	0.1308(0.0693)	0.1538(0.0643)
0.99	$\hat{g}_{LS}(x)$	0.0218(0.0367)	0.0386(0.0521)	0.0684(0.0628)
	$\hat{g}_{j1}(x)$	0.7654(0.2416)	0.9902(0.2863)	0.2495(0.2813)
	$\hat{g}_{S}(x)$	0.2062(0.0585)	0.2484(0.0633)	0.2954(0.0603)

Table 4 MSEs (Standard Deviation) based on 500 replications and n=200, GLD(0,7,7,7)

		h(n)		
x	estimate	0.15	0.19	0.25
0.01	$\hat{g}_{LS}(x)$	0.0581(0.0425)	0.0741(0.0391)	0.1084(0.0375)
	$\hat{g}_{j1}(x)$	0.3120(0.0349)	0.1223(0.0337)	0.1648(0.0229)
	$\hat{g}_S(x)$	0.4556(0.0339)	0.5246(0.0189)	0.5338(0.0116)
0.1045	$\hat{g}_{LS}(x)$	0.0021(0.0029)	0.0026(0.0031)	0.0048(0.0041)
	$\hat{g}_{j1}(x)$	0.0524(0.0125)	0.0426(0.0087)	0.0261(0.0054)
	$\hat{g}_S(x)$	0.0408(0.0099)	0.0492(0.0068)	0.0480(0.0043)
0.2040	$\hat{g}_{LS}(x)$	0.0022(0.0015)	0.0035(0.0017)	0.0045(0.0020)
	$\hat{g}_{j1}(x)$	0.0014(0.0007)	0.0001(0.0001)	0.0030(0.0009)
	$\hat{g}_S(x)$	0.0008(0.0006)	0.0003(0.0003)	0.0044(0.0011)
0.4030	$\hat{g}_{LS}(x)$	0.0028(0.0014)	0.0068(0.0019)	0.0151(0.0025)
	$\hat{g}_{j1}(x)$	0.0218(0.0014)	0.0340(0.0016)	0.0587(0.0019)
	$\hat{g}_S(x)$	0.0220(0.0015)	0.0344(0.0017)	0.0596(0.0020)
0.6020	$\hat{g}_{LS}(x)$	0.0038(0.0017)	0.0081(0.0022)	0.0165(0.0024)
	$\hat{g}_{j1}(x)$	0.0212(0.0014)	0.0333(0.0016)	0.0578(0.0018)
	$\hat{g}_S(x)$	0.0214(0.0015)	0.0337(0.0017)	0.0587(0.0020)
0.8010	$\hat{g}_{LS}(x)$	0.0039(0.0018)	0.0049(0.0017)	0.0053(0.0018)
	$\hat{g}_{j1}(x)$	0.0020(0.0009)	0.0002(0.0002)	0.0021(0.0007)
	$\hat{g}_S(x)$	0.0013(0.008)	0.0001(0.0002)	0.0032(0.0009)
0.9005	$\hat{g}_{LS}(x)$	0.0023(0.0034)	0.0031(0.0039)	0.0055(0.0044)
	$\hat{g}_{j1}(x)$	0.0563(0.0141)	0.0476(0.0102)	0.0304(0.0058)
	$\hat{g}_S(x)$	0.0478(0.0108)	0.0572(0.0076)	0.0558(0.0046)
0.99	$\hat{g}_{LS}(x)$	0.782(0.0534)	0.0877(0.0464)	0.1155(0.0370)
	$\hat{g}_{j1}(x)$	0.0324(0.0374)	0.1275(0.0376)	0.1667(0.0240)
	$\hat{g}_S(x)$	0.4416(0.0365)	0.5134(0.0200)	0.5240(0.0121)

5. Proofs

5.1. Auxiliary results

The following lemmas show several statistical properties (moments and concentration inequalities) of the estimators $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ given in (2.4) and (2.5), respectively. The proofs or appropriate references are given in the next sub-section.

Lemma 5.1. *Let $p \geq 1$, $\hat{c}_{j,k}$ be defined by* (2.4) *and $\hat{d}_{j,k}$ be defined by* (2.5)*. Then*

- *there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{c}_{j,k} - c_{j,k})^{2p}) \leq C 2^{3jp} \frac{1}{n^p}.$$

- *there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{d}_{j,k} - d_{j,k})^{2p}) \leq C 2^{3jp} \frac{1}{n^p}.$$

Lemma 5.2. *Let $p \geq 1$, $\hat{d}_{j,k}$ be defined by* (2.5) *and λ_j be defined by* (3.1)*. Then*

$$\mathbb{P}\left(|\hat{d}_{j,k} - d_{j,k}| \geq \lambda_j\right) \leq 2 \frac{1}{n^p}.$$

Lemma 5.3. *Suppose that* **(A)** *is satisfied. Let $p \geq 1$, j such that $2^j \leq \sqrt{n/\ln(n)}$, $\hat{c}_{j,k}$ be defined by* (2.4) *and $\hat{d}_{j,k}$ be defined by* (2.5)*. Then*

- *there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{c}_{j,k} - c_{j,k})^{2p}) \leq C \frac{1}{n^p}.$$

- *there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{d}_{j,k} - d_{j,k})^{2p}) \leq C \frac{1}{n^p}.$$

Lemma 5.4. *Suppose that* **(A)** *is satisfied. Let $p \geq 1$, j such that $2^j \leq \sqrt{n/\ln(n)}$, $\hat{d}_{j,k}$ be defined by* (2.5) *and λ_j be defined by* (3.2)*. Then*

$$\mathbb{P}\left(|\hat{d}_{j,k} - d_{j,k}| \geq \lambda_j\right) \leq 2 \frac{1}{n^p}.$$

In the rest of the paper, we use C to denote positive constants whose value may change from line to line.

5.2. Proofs of the auxiliary results

Proof of Lemma 5.1. Note that $(\phi_{j,k}(x))' = 2^{3j/2}\phi'(2^j x - k)$ and $K = \sup_{x\in[0,1]} |\phi'(x)| < \infty$. The mean value theorem gives, for any $(x, y) \in [0, 1]^2$,

$$|\phi_{j,k}(x) - \phi_{j,k}(y)| \leq 2^{3j/2} K|x - y|.$$

Therefore

$$\begin{aligned} |\hat{c}_{j,k} - c_{j,k}| &= \left| \int_{[0,1]} \left(\phi_{j,k}(\hat{F}(x)) - \phi_{j,k}(F(x)) \right) dx \right| \\ &\leq \int_{[0,1]} |\phi_{j,k}(\hat{F}(x)) - \phi_{j,k}(F(x))| dx \\ &\leq C 2^{3j/2} \int_{[0,1]} |\hat{F}(x) - F(x)| dx. \end{aligned} \tag{5.1}$$

Using the Hölder inequality and the Fubini theorem, we have

$$\begin{aligned} \mathbb{E}((\hat{c}_{j,k} - c_{j,k})^{2p}) &\leq C 2^{3jp} \mathbb{E}\left(\left(\int_{[0,1]} |\hat{F}(x) - F(x)| dx \right)^{2p} \right) \\ &\leq C 2^{3jp} \left(\int_{[0,1]} \mathbb{E}\left((\hat{F}(x) - F(x))^{2p} \right) dx \right). \end{aligned}$$

The Rosenthal inequality (see [38]) yields

$$\begin{aligned} &\mathbb{E}\left((\hat{F}(x)) - F(x))^{2p} \right) \\ &\leq C \left(\frac{(1 - F(x))^{2p} F(x) + (F(x))^{2p}(1 - F(x))}{n^{2p-1}} + \frac{(F(x)(1 - F(x)))^p}{n^p} \right) \\ &\leq C \frac{1}{n^p}. \end{aligned}$$

Hence

$$\mathbb{E}((\hat{c}_{j,k} - c_{j,k})^{2p}) \leq C 2^{3jp} \frac{1}{n^p}.$$

Working with ψ instead of ϕ, we show that

$$\mathbb{E}((\hat{d}_{j,k} - d_{j,k})^{2p}) \leq C 2^{3jp} \frac{1}{n^p}.$$

The proof of Lemma 5.1 is complete. □

Proof of Lemma 5.2. By (5.1) with ψ instead of ϕ, we obtain

$$\begin{aligned} |\hat{d}_{j,k} - d_{j,k}| &\leq K 2^{3j/2} \int_{[0,1]} |\hat{F}(x) - F(x)| dx \\ &\leq K 2^{3j/2} \sup_{x\in[0,1]} |\hat{F}(x) - F(x)|, \end{aligned}$$

with $K = \sup_{x\in[0,1]} |\psi'(x)|$. It follows from the Massart version of the Dvoretzky-Kiefer-Wolfowitz inequality (see [15] and [28]) and the definition of λ_j (3.1) that

$$\begin{aligned} &\mathbb{P}\left(|\hat{d}_{j,k} - d_{j,k}| \geq \frac{\lambda_j}{2} \right) \\ &\leq \mathbb{P}\left(K 2^{3j/2} \sup_{x\in[0,1]} |\hat{F}(x) - F(x)| \geq \frac{1}{2} K 2^{3j/2} \sqrt{2p} \sqrt{\frac{\ln(n)}{n}} \right) \\ &\leq \mathbb{P}\left(\sup_{x\in[0,1]} |\hat{F}(x) - F(x)| \geq \sqrt{\frac{p}{2}} \sqrt{\frac{\ln(n)}{n}} \right) \\ &\leq 2\exp\left(-2n \left(\sqrt{\frac{p}{2}} \sqrt{\frac{\ln(n)}{n}} \right)^2 \right) = 2\frac{1}{n^p}. \end{aligned}$$

This ends the proof of Lemma 5.2. □

Proof of Lemmas 5.3 and 5.4. By the change of variable $x = Q(y)$, let us observe that

$$\begin{aligned} \hat{d}_{j,k} - d_{j,k} &= \int_{[0,1]} \left(\psi_{j,k}(\hat{F}(x)) - \psi_{j,k}(F(x)) \right) dx \\ &= \int_{[0,1]} \left(\psi_{j,k}(\hat{U}(x)) - \psi_{j,k}(x) \right) g(x) dx, \end{aligned}$$

with

$$\hat{U}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{U_i \leq x\}}, \qquad U_i = F(X_i).$$

Then the proofs of Lemmas 5.3 and 5.4 follow from the technical part of [26, Subsection 9.2.2. pages 1093 - 1098] with g instead of $f(G^{-1})$. Let us mention that for the validity of results we need to suppose **(A)** and a restriction on j is considered in our study, i.e., $2^j \leq \sqrt{n/\ln(n)}$. □

5.3. Proofs of the main results

Proof of Theorem 3.1. We expand g on the wavelet basis $\mathcal{W}$ as

$$g(x) = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} d_{j,k} \psi_{j,k}(x),$$

where $c_{j_0,k} = \int_{[0,1]} g(x)\phi_{j_0,k}(x)dx$ and $d_{j,k} = \int_{[0,1]} g(x)\psi_{j,k}(x)dx$. Hence

$$\hat{g}_L(x) - g(x) = \sum_{k=0}^{2^{j_0}-1} (\hat{c}_{j_0,k} - c_{j_0,k})\phi_{j_0,k}(x) - \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x).$$

Using the inequality : $|x+y|^p \leq 2^{p-1}(|x|^p + |y|^p)$, $(x,y) \in R^2$, we obtain

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_L(x) - g(x)|^p dx\right) \leq 2^{p-1}(F+G), \tag{5.2}$$

where

$$F = \mathbb{E}\left(\int_{[0,1]} \left|\sum_{k=0}^{2^{j_0}-1} (\hat{c}_{j_0,k} - c_{j_0,k})\phi_{j_0,k}(x)\right|^p dx\right)$$

and

$$G = \int_{[0,1]} \left|\sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(x)\right|^p dx.$$

Using a $\mathbb{L}_p$ norm result on wavelet series (see [21, Proposition 8.3]), the Cauchy-Schwarz inequality and Lemma 5.1, we obtain

$$\begin{aligned} F &\leq C2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left(|\hat{c}_{j_0,k} - c_{j_0,k}|^p\right) \\ &\leq C2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \left(\mathbb{E}\left((\hat{c}_{j_0,k} - c_{j_0,k})^{2p}\right)\right)^{1/2} \\ &\leq C2^{j_0(p/2-1)} 2^{j_0} 2^{3j_0 p/2} \frac{1}{n^{p/2}} = C\left(\frac{2^{4j_0}}{n}\right)^{p/2}. \end{aligned} \tag{5.3}$$

On the other hand, using $g \in B^s_{r,q}(M)$ and proceeding as in [14, eq. (24)], we have immediately

$$G \leq C2^{-j_0 s_* p}. \tag{5.4}$$

It follows from (5.2), (5.3), (5.4) and the definition of j_0 that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{g}_L(x) - g(x)|^p dx\right) \leq C\left(\left(\frac{2^{4j_0}}{n}\right)^{p/2} + 2^{-j_0 s_* p}\right) \leq Cn^{-s_* p/(2s_*+4)}.$$

This ends the proof of Theorem 3.1.

□

Proof of Theorem 3.2. Theorem 3.2 is a consequence of Theorem 5.1 below with $\nu = 3/2$ and Lemmas 5.1 and 5.2 above.

Theorem 5.1. *Let $h \in \mathbb{L}_2([0,1])$ be an unknown function to be estimated from n observations and consider its wavelet decomposition. Let $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ be estimators of these wavelet coefficients $c_{j,k}$ and $d_{j,k}$ respectively. We suppose that there exist three constants $\nu > 0$, $C > 0$ and $\kappa > 0$ such that the following inequalities hold :*

Moments inequalities : *for any $j \geq \tau$ such that $2^j \leq n$ and $k \in \{0, \ldots, 2^j - 1\}$,*

$$\mathbb{E}\left((\hat{c}_{j,k} - c_{j,k})^{2p}\right) \leq C 2^{2\nu jp} \left(\frac{\ln(n)}{n}\right)^p$$

and

$$\mathbb{E}\left((\hat{d}_{j,k} - d_{j,k})^{2p}\right) \leq C 2^{2\nu jp} \left(\frac{\ln(n)}{n}\right)^p.$$

Concentration inequality : *for any $j \geq \tau$ such that $2^j \leq n/\ln(n)$ and $k \in \{0, \ldots, 2^j - 1\}$,*

$$\mathbb{P}\left(|\hat{d}_{j,k} - d_{j,k}| \geq \frac{\kappa}{2} 2^{\nu j} \sqrt{\frac{\ln(n)}{n}}\right) \leq C \left(\frac{\ln(n)}{n}\right)^p.$$

Let us define the hard thresholding wavelet estimator of h by

$$\hat{h}(x) = \sum_{k=0}^{2^\tau - 1} \hat{c}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k=0}^{2^j - 1} \hat{d}_{j,k} \mathbf{1}_{\left\{|\hat{d}_{j,k}| \geq \kappa 2^{\nu j} \sqrt{\frac{\ln(n)}{n}}\right\}} \psi_{j,k}(x), \qquad x \in [0,1],$$

where j_1 is the integer satisfying

$$\left(\frac{n}{\ln(n)}\right)^{1/(2\nu+1)} < 2^{j_1+1} \leq 2 \left(\frac{n}{\ln(n)}\right)^{1/(2\nu+1)}.$$

Suppose that $h \in B^s_{r,q}(M)$ with $s > 0$, $r \geq 1$ and $q \geq 1$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}\left(\int_{[0,1]} |\hat{h}(x) - h(x)|^p dx\right) \leq C \Theta_{n,\nu},$$

where

$$\Theta_{n,\nu} = \begin{cases} \left(\frac{\ln(n)}{n}\right)^{sp/(2s+2\nu+1)}, & \text{for} \quad rs > (\nu + 1/2)(p-r), \\ \left(\frac{\ln(n)}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2\nu+1)}, & \text{for} \quad rs < (\nu + 1/2)(p-r), \\ \left(\frac{\ln(n)}{n}\right)^{(s-1/r+1/p)p/(2s-2/r+2\nu+1)} (\ln(n))^{(p-r/q)_+}, & \text{for} \quad rs = (\nu + 1/2)(p-r). \end{cases}$$

The general form of Theorem 5.1 can be proved using arguments similar to Theorem 5.1 of [25] for a bound of the $\mathbb{L}_p$-risk and Theorem 4.2 of [8] for the determination of the rates of convergence.

□

Proof of Theorem 3.3. The proof follows step by step the one of Theorem 3.1 with the use of Lemma 5.3 instead of Lemma 5.1. Thanks to the new definition of j_0, we finally obtain

$$\mathbb{E}\left(\int_{[0,1]}|\hat{g}_L(x)-g(x)|^p dx\right) \leq C\left(\left(\frac{2^{j_0}}{n}\right)^{p/2}+2^{-j_0 s_* p}\right) \leq Cn^{-s_* p/(2s_*+1)}.$$

□

Proof of Theorem 3.4. The proof follows along the lines of that of Theorem 3.2 with the use of Lemmas 5.3 and 5.4 instead of Lemmas 5.1 and 5.2, and an adjustment on the rates of convergence as in [26, Proposition 2]. This minor modification is due to our definition of j_1.

□

ACKNOWLEDGEMENTS - We thank the associate editor and the referees for their critical comments which have led to the improved version of the paper.

References

[1] Abbaszadeh, M., Chesneau, C. and Doosti, H. (2012). Nonparametric estimation of a density under bias and multiplicative censoring via wavelet methods, *Statistics and Probability Letters,* **82**, 932- 941.

[2] Abbaszadeh, M., Chesneau, C. and Doosti, H. (2013). Multiplicative censoring : estimation of a density and its derivatives under the L_p-risk, *Revstat,* **11**, 255-276.

[3] Antoniadis, A. (1997). Wavelets in statistics : a review (with discussion), *Journal of the Italian Statistical Society Series B*, **6**, 97-144.

[4] Caroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density, *Journal American Statistical Association*, **83**, 1184-1186.

[5] Chagny, G. (2013a). Penalization versus Goldenshluger-Lepski strategies in regression estimation with warped bases, *Esaim Probability and Statistics*, **17**, 328-358.

[6] Chagny, G. (2013b). Warped bases for conditional density estimation, *Mathematical Methods of Statistics*, **22**, (4), 253-282.

[7] Chagny, G. and Lacour, C. (2015). Optimal adaptive estimation of the relative density, *TEST (DOI : 10.1007/s11749-015-0426-6)*, to appear.

[8] Chesneau, C. (2008). Wavelet estimation via block thresholding : a minimax study under $\mathbb{L}^p$ risk, *Statistica Sinica*, **18**, 3, 1007-1024.

[9] Chesneau, C., Dewan, I and Doosti, H. (2012). Wavelet linear density estimation for associated stratified size-biased sample. *Journal of Nonparametric Statistics,* **24**, (2), 429-445.

[10] Chesneau, C., Dewan, I. and Doosti, H. (2014). Nonparametric estimation of two dimensional continuous- discrete density function by wavelets. *Statistical Methodology,* **18**, 64-78.

[11] Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms, *Applied and Computational Harmonic Analysis*, **24**, 1, 54-81.

[12] Daubechies, I. (1992). *Ten lectures on wavelets*, SIAM.

[13] Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Applied Computational Harmonic Analysis*, **3**, 215-228.

[14] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding, *Annals of Statistics*, **24**, 508-539.

[15] Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Annals Mathematical Statistics*, **27**, 642-669.

[16] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem. *Annals Statistics*, **19**, 1257-1272.

[17] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of American Statistical Association,* **87**, 998-1004.

[18] Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Annals of Statistics*, **21**, 196-216.

[19] Freimer, M., Kollia, G., Mudholkar, G.S. and Lin, C.T. (1988). A study of the Generalized Tukey Lambda family, *Communications in Statistics - Theory and Methods,* **17**, 3547-3567.

[20] Gilchrist,W. (2000). *Statistical Modeling with Quantile Functions*, Chapman and Hall, New York.

[21] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, Springer Verlag. New York.

[22] Jeong, J.H. and Fine, J.P. (2009). Parametric regression on cumulative incidence function, *Biostatistics.* **8**, 184-196.

[23] Jones, M.C. (1992). Estimating densities, quantiles, quantile densities and density quantiles, *Annals Institute Statistical Mathematics* **44**, 4, 721-727.

[24] Karian, Z. A. and Dudewicz, E. J. (2000). *Fitting statistical distributions : The Generalized lambda distribution and generalized bootstrap methods*, CRC, London.

[25] Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, **9**, 283-345.

[26] Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets, *Bernoulli*, **10**, 1053-1105.

[27] Mallat, S. (2009). *A wavelet tour of signal processing. The sparse way.* With contributions from Gabriel Peyré, Elsevier/ Academic Press, Amsterdam.

[28] Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, *The Annals of Probability*, 18, 1269-1283.

[29] Meyer, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.

[30] Nair, N.U., Sankaran, P.G. and Kumar, B.V. (2008). Total time on test transforms of order n and its implications in reliability analysis, *Journal of Applied Probability*, **45**, 1126-1139.

[31] Nair, N.U. and Sankaran, P.G. (2009). Quantile based reliability analysis, *Communications Statistics : Theory Methods.* **38**, 222-232.

[32] Parzen, E. (1979). Non parametric statistical data modeling, *Journal of American Statistical Association,* **74**, 105-122.

[33] Parzen, E. (2004). Quantile Probability and Statistical Data Modeling, *Statistical Science*, **19**, 4, 652-662.

[34] Peng, L. and Fine, J.P. (2007). Nonparametric quantile inference with competing risks data, *Biometrika*, **94**, 735-744.

[35] Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*, Academic Press, Orlando.

[36] Ramirez, P. and Vidakovic, B. (2010). Wavelet density estimation for stratified size-biased sample. *Journal of Statistical planning and Inference,* **140**, 419-432.

[37] Reid, N. (1981). Estimating the median survival time, *Biometrika*, **68**, 601-608.

[38] Rosenthal, H.P. (1970). On the subspaces of L_p $(p \geq 2)$ spanned by sequences of independent random variables, *Israel Journal of Mathematics*, **8**, 273-303.

[39] Sankaran, P.G. and Nair N.U. (2009). Nonparametric estimation of hazard quantile function, *Journal of Nonparametric Statistics*, **21**, 757-767.

[40] Sankaran, P.G., Nair, N.U. and Sreedevi, E.P. (2010). A quantile based test for comparing cumulative incidence functions of competing risks models, *Statist. Probab. Lett.*, **80**, 886-891.

[41] Sankaran, P.G., Dewan, I. and Sreedevi E.P. (2015). A non-parametric test for stochastic dominance using total time on test transform, *American Journal of Mathematical and Management Sciences*, **34**, 162-183.

[42] Shirazi, E., Chaubey, Y. P., Doosti, H. and Niroumand, H. (2012). Wavelet based estimation for the derivative of a density by block thresholding under random censorship, *Journal of the Korean Statistical Society,* **41**, 199-211.

[43] Shirazi, E., Doosti, H., Niroumand, H.A. and Hosseinioun, N. (2013). Nonparametric regression estimates with censored data based on block thresholding method, *Journal of Statistical Planning and Inference*, **143**, 1150-1165.

[44] Slud, E.V., Byar, D.P. and Green, S.B. (1984). A comparison of reflected versus testbased confidence intervals for the median survival time, based on censored data, *Biometrics*, **40**, 587-600.

[45] Soni, P., Dewan, I. and Jain, K. (2012). Nonparametric estimation of quantile density function, *Computational Statistics and Data Analysis*, **56**, 12, 3876-3886.

[46] Soni, P., Dewan, I. and Jain, K. (2015a). Tests for successive differences of quantiles, *Statistics and Probability Letters*, **97**, 1-8.

[47] Soni, P., Dewan, I. and Jain, K. (2015b). Nonparametric multiple comparison methods for quantiles, *Submitted*.

[48] Su, J.Q. and Wei, L.J. (1993). Nonparametric estimation for the difference or ratio of median failure times, *Biometrics*, **49**, 603-607.

[49] Tsybakov, A. (2004). *Introduction à l'estimation nonparamétrique*, Springer Verlag, Berlin.

[50] Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, John Wiley & Sons, Inc., New York, 384 pp.