REVSTAT – Statistical Journal
Volume 11, Number 3, November 2013, 255–276

MULTIPLICATIVE CENSORING: ESTIMATION OF A DENSITY AND ITS DERIVATIVES UNDER THE L_p-RISK

Authors: Mohammad Abbaszadeh
– Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran
abbaszadeh.mo@stu-mail.um.ac.ir

Christophe Chesneau
– Université de Caen, LMNO, Campus II, Science 3, 14032, Caen, France
christophe.chesneau@unicaen.fr

Hassan Doosti
– Department of Mathematics, Kharazmi University, Tehran, Iran
and
Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
hassan.doosti@unimelb.edu.au

Received: May 2012 Revised: January 2013 Accepted: January 2013

Abstract:

- We consider the problem of estimating a density and its derivatives for a sample of multiplicatively censored random variables. The purpose of this paper is to present an approach to this problem based on wavelets methods. Two different estimators are developed: a linear based on projections and a nonlinear using a term-by-term selection of the estimated wavelet coefficients. We explore their performances under the L_p-risk with $p \geq 1$ and over a wide class of functions: the Besov balls. Fast rates of convergence are obtained. Finite sample properties of the estimation procedure are studied on a simulated data example.

Key-Words:

- *density estimation; multiplicative censoring; inverse problem; wavelets; Besov balls; L_p-risk.*

AMS Subject Classification:

- 62G07, 62G20.

1. INTRODUCTION

The multiplicative censoring density model can be described as follows. We observe n *i.i.d.* random variables $Y_1, ..., Y_n$ where, for any $i \in \{1, ..., n\}$,

$$Y_i = U_i X_i \ , \tag{1.1}$$

$U_1, ..., U_n$ are n unobserved *i.i.d.* random variables having the common uniform distribution on $[0, 1]$ and $X_1, ..., X_n$ are n unobserved *i.i.d.* random variables with common unknown density $f : [0, 1] \to [0, \infty]$. For any $i \in \{1, ..., n\}$, we suppose that U_i and X_i are independent. Our aim is to estimate f (or a transformation of f) from $Y_1, ..., Y_n$. Details, applications and results of this model can be found in, e.g., [37], [38], [2] and [1]. For recent applications in the field of signal processing, we refer to [7] and references therein for further readings.

In this paper, we investigate the estimation of $f^{(m)}$ (including f for $m = 0$). This is particularly of interest to detect possible bumps, concavity or convexity properties of f. The estimation of the derivatives of a density have been investigated by several authors. The pioneers are [4], [35] and [36]. Recent studies can be found in [31], [9, 10], [32] and [8].

In recent years, wavelet methods in nonparametric function estimation have become a powerful technique. The major advantages of these methods are their spatial adaptivity and asymptotic optimality properties over large function spaces. We refer to, e.g., [3], [23] and [39]. These facts motivate the estimation of $f^{(m)}$ via wavelet methods. To the best of our knowledge, this has never been investigated before for (1.1). Combning the approaches of [1] and [31], we construct two different wavelet estimators: a linear one and a nonlinear adaptive one based on a hard thresholding rule introduced by [18]. The latter method has the advantage to be adaptive; it does not depend on the knowledge of the smoothness of $f^{(m)}$ in its construction. We explore their performances via the L_p-risk with $p \geq 1$ (including the Mean Integrated Squared Error (MISE) which corresponds to $p = 2$) over a "standard" wide class of unknown functions: the Besov balls $B^s_{r,q}(M)$. Our main result proves that the considered adaptive wavelet estimator achieves a fast rate of convergence. Then we show the finite sample properties of the considered estimators by a simulated data.

The rest of the paper is organized as follows. Section 2 briefly describes the wavelet basis and the Besov balls. Assumptions on the model and the wavelet estimators are presented in Section 3. The theoretical results are given in Section 4. A simulation study is done in Section 5. The proofs are gathered in Section 6.

2. WAVELETS AND BESOV BALLS

This section is devoted to bascics on wavelets and Besov balls.

2.1. Wavelets

Let N be a positive integer such that $N > 10(m+1)$ (where m refers to the estimation of $f^{(m)}$).

Throughout the paper, we work within an orthonormal multiresolution analysis of $L_2([0,1]) = \{h\colon [0,1] \to \mathbb{R};\ \int_0^1 h(x)^2\,dx < \infty\}$, associated with the initial wavelet functions ϕ and ψ of the Daubechies wavelets $db2N$. The features of these functions are to be compactly supported and $\mathcal{C}^{m+1}$.

Set

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

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

$$\mathcal{S} = \Big\{\phi_{\ell,k};\ k \in \{0,...,2^\ell-1\};\ \psi_{j,k};\ j \in \mathbb{N} - \{0,...,\ell-1\},\ k \in \{0,...,2^j-1\}\Big\}$$

is an orthonormal basis of $L_2([0,1])$.

For any integer $\ell \geq \tau$, any $h \in L_2([0,1])$ can be expanded on $\mathcal{S}$ as

$$h(x) = \sum_{k=0}^{2^\ell-1} c_{\ell,k}\,\phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\,\psi_{j,k}(x)\ , \qquad x \in [0,1]\ , \tag{2.1}$$

where

$$c_{j,k} = \int_0^1 h(x)\,\phi_{j,k}(x)\,dx\ , \qquad d_{j,k} = \int_0^1 h(x)\,\psi_{j,k}(x)\,dx\ . \tag{2.2}$$

See, e.g., [14] and [27].

As usual in nonparametric statistics via wavelets, we will suppose that the unknown function $f^{(m)}$ belongs to Besov balls defined below.

2.2. Besov balls

Let $M > 0$, $s > 0$, $r \geq 1$, $q \geq 1$ and $L_r([0,1]) = \{h\colon [0,1] \to \mathbb{R};\ \int_0^1 |h(x)|^r dx < \infty\}$. Set, for every measurable function h on [0,1] and $\epsilon \geq 0$, $\Delta_\epsilon(h)(x) = h(x+\epsilon) - h(x)$, $\Delta_\epsilon^2(h)(x) = \Delta_\epsilon(\Delta_\epsilon(h))(x)$ and, identically, $\Delta_\epsilon^N(h)(x) = \Delta_\epsilon^{N-1}(\Delta_\epsilon(h))(x)$.

Let

$$\rho^N(t,h,r) \;=\; \sup_{|\epsilon|\le t} \left(\int_0^1 |\Delta_\epsilon^N(h)(x)|^r \, dx \right)^{1/r} .$$

Then, for $s \in [0,N)$, we define the Besov ball $B^s_{r,q}(M)$ by

$$B^s_{r,q}(M) \;=\; \left\{ h \in L_r([0,1]) ;\; \left(\int_0^1 \left(\frac{\rho^N(t,h,r)}{t^s} \right)^q \frac{dt}{t} \right)^{1/q} \le M \right\} ,$$

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

We have the equivalence: $h \in B^s_{r,q}(M)$ if and only if there exists a constant $M^* > 0$ (depending on M) such that (2.2) satisfy

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

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 [28] and [23, Chapter 9].

3. ESTIMATORS

This section describes our wavelet estimation approach.

3.1. Wavelet methodology

Suppose that, for any $v \in \{0,...,m\}$, $f^{(v)} \in L_2([0,1])$. Then we have the wavelet series expansion:

$$f^{(m)}(x) \;=\; \sum_{k=0}^{2^\ell-1} c^{(m)}_{\ell,k} \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k=0}^{2^j-1} d^{(m)}_{j,k} \psi_{j,k}(x) \;, \qquad x \in [0,1] \;,$$

where $c^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\,\phi_{j,k}(x)\,dx$ and $d^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\,\psi_{j,k}(x)\,dx$ and m is the order of the density derivative to be estimated.

We now aim to construct natural estimators for these unknown wavelet coefficients. Combining the approaches of [1] and [31], let us investigate a more arranging expression for $c^{(m)}_{j,k}$ (the same development holds for $d^{(m)}_{j,k}$).

Suppose that, for any $v \in \{0, ..., m\}$, $f^{(v)}(0) = f^{(v)}(1) = 0$. It follows from m-fold integration by parts that

$$c_{j,k}^{(m)} = (-1)^m \int_0^1 f(x)\,(\phi_{j,k})^{(m)}(x)\,dx\ .$$

Note that, since $U_1 \sim \mathcal{U}([0,1])$ and U_1 and X_1 are independent, the density of Y_1 is

$$g(x) = \int_x^1 \frac{f(y)}{y}\,dy\ , \qquad x \in [0,1]\ . \tag{3.1}$$

Hence $f(x) = -xg'(x)$, $x \in [0,1]$.

One integration by parts yields

$$\begin{aligned} c_{j,k}^{(m)} &= (-1)^m \left(-\int_0^1 g'(x)\,x(\phi_{j,k})^{(m)}(x)\,dx\right) \\ &= (-1)^m \int_0^1 \Big((\phi_{j,k})^{(m)}(x) + x(\phi_{j,k})^{(m+1)}(x)\Big)\,g(x)\,dx \\ &= \mathbb{E}\bigg((-1)^m \Big((\phi_{j,k})^{(m)}(Y_1) + Y_1(\phi_{j,k})^{(m+1)}(Y_1)\Big)\bigg)\ . \end{aligned}$$

The method of moments gives the following unbiased estimator for $c_{j,k}^{(m)}$:

$$\hat{c}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^{n} \Big((\phi_{j,k})^{(m)}(Y_i) + Y_i(\phi_{j,k})^{(m+1)}(Y_i)\Big) \tag{3.2}$$

and, similarly, an unbiased estimator for $d_{j,k}^{(m)}$ is

$$\hat{d}_{j,k}^{(m)} = \frac{(-1)^m}{n} \sum_{i=1}^{n} \Big((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i)\Big)\ . \tag{3.3}$$

Further properties of these wavelet coefficients estimators are explored in Propositions 6.1 and 6.2 below. We are now in the position to present the considered estimators for $f^{(m)}$.

3.2. Main estimators

We define the linear estimator $\hat{f}_{lin}^{(m)}$ by

$$\hat{f}_{lin}^{(m)}(x) = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k}^{(m)}\,\phi_{j_0,k}(x)\ , \qquad x \in [0,1]\ , \tag{3.4}$$

where $\hat{c}_{j,k}^{(m)}$ is defined by (3.2) and j_0 is an integer which will be properly chosen later.

Recent developments on the linear wavelet estimators for various density estimation problems can be found in [11].

We define the hard thresholding estimator $\hat{f}^{(m)}_{hard}$ by

$$\hat{f}^{(m)}_{hard}(x) = \sum_{k=0}^{2^\tau-1} \hat{c}^{(m)}_{\tau,k}\,\phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1}\sum_{k=0}^{2^j-1} \hat{d}^{(m)}_{j,k}\,\mathbf{1}_{\left\{|\hat{d}^{(m)}_{j,k}|\geq\kappa\delta^{(m)}_j\right\}}\,\psi_{j,k}(x) , \qquad x\in[0,1] , \tag{3.5}$$

where $\hat{c}^{(m)}_{j,k}$ and $\hat{d}^{(m)}_{j,k}$ are defined by (3.2) and (3.3), $\mathbf{1}$ is the indicator function, j_1 is the integer satisfying

$$\left(\frac{n}{\ln n}\right)^{1/(2m+3)} < 2^{j_1+1} \leq 2\left(\frac{n}{\ln n}\right)^{1/(2m+3)} ,$$

$\delta^{(m)}_j$ is the threshold:

$$\delta^{(m)}_j = 2^{j(m+1)}\sqrt{\frac{\ln n}{n}}$$

and κ is a large enough constant (see Remark 4.2 and Proposition 6.2).

The major difference between $\hat{f}^{(m)}_{lin}$ and $\hat{f}^{(m)}_{hard}$ is the term-by-term selection of the wavelet coefficients estimators which makes $\hat{f}^{(m)}_{hard}$ adaptive. Discussions on hard thresholding estimators in nonparametric function estimation can be found in, e.g., [18], [23], [16] and [39].

Remark 3.1. A preliminary idea is to rewrite the model (1.1) as: $-\ln Y_i = -\ln X_i - \ln U_i$. In this form, it becomes the standard density deconvolution model where $-\ln U_1, ..., -\ln U_n$ are n unobserved *i.i.d.* random variables having the common exponential distribution with parameter 1 and $-\ln X_1, ..., -\ln X_n$ are n unobserved *i.i.d.* random variables with unknown density

$$q(x) = e^{-x} f(e^{-x}) , \qquad x\in(0,\infty) .$$

Then there exist a wide variety of methods to estimate q. See, e.g., [19], [22], [29], [5], [15] and [26]. Results on the estimation of $q^{(m)}$ via kernel methods can be found in [19]. However, due to the definition of q, it seems difficult to deduce results on the estimation of $f^{(m)}$ from $q^{(m)}$ under the L_p-risk.

Remark 3.2. Another possible approach to estimate $f^{(m)}$ is described below. Since $f(x) = -xg'(x)$, $x\in[0,1]$, we have

$$f^{(m)}(x) = -\big(mg^{(m)}(x) + xg^{(m+1)}(x)\big) , \qquad x\in[0,1] . \tag{3.6}$$

Then a plug-in approach to estimate $f^{(m)}$ consists in estimating $g^{(m)}$ by $\hat{g}^{(m)}$ and $g^{(m+1)}$ by $\hat{g}^{(m+1)}$, and to inject them in (3.6). This yields the estimator

$$\hat{f}^{(m)}_*(x) = -\big(m\hat{g}^{(m)}(x) + x\hat{g}^{(m+1)}(x)\big) , \qquad x\in[0,1] .$$

However, there are at least two disadvantages to this approach.

- Firstly, since two different estimators are required, more errors are injected in $\hat{f}_*^{(m)}$ in comparison to (3.5).
- Secondly, the choices of $\hat{g}^{(m)}$ and $\hat{g}^{(m+1)}$ are not so clear. If we focus our attention on wavelet estimators, one can chose hard thresholding versions as in [31]. However, the presence of x in front of $\hat{g}^{(m+1)}(x)$ implies that we work with the nonorthonormal basis $\mathcal{S}_* = \{x\phi_{\ell,k}(x),\ k \in \{0, ..., 2^\ell - 1\};\ x\psi_{j,k}(x);\ j \in \mathbb{N} - \{0, ..., \ell - 1\},\ k \in \{0, ..., 2^j - 1\}\}$. And it is not immediately clear how we can manipulate it in the context of the L_p-risk.

4. RESULTS

Before presenting the main results, let us formulate the following assumptions:

(**A1**) for any $v \in \{0, ..., m\}$, $f^{(v)}(0) = f^{(v)}(1) = 0$,

(**A2**) there exists a known constant $C > 0$ such that, for any $v \in \{0, ..., m\}$,

$$\int_0^1 \left(f^{(v)}(x)\right)^2 dx \le C ,$$

(**A3**) there exists a known constant $C > 0$ such that

$$\sup_{x \in [0,1]} g(x) \le C ,$$

where g is as in (3.1).

Theorems 4.1 and 4.2 below explore the performance of our estimators under the L_p-risk over Besov balls.

Theorem 4.1 (L_p-risk for $\hat{f}_{lin}^{(m)}$). *Consider* (1.1) *under (A1), (A2) and (A3). Let* $p \ge 1$. *Suppose that* $f^{(m)} \in B^s_{r,q}(M)$ *with* $s > 0$, $r \ge 1$ *and* $q \ge 1$. *Set* $s_* = \min(s, s - 1/r + 1/p)$ *and let* $\hat{f}_{lin}^{(m)}$ *be as in* (3.4) *with* j_0 *being the integer such that*

$$n^{1/(2s_*+2m+3)} < 2^{j_0+1} \le 2n^{1/(2s_*+2m+3)} .$$

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

$$\mathbb{E}\left(\int_0^1 \left(\hat{f}_{lin}^{(m)}(x) - f^{(m)}(x)\right)^p dx\right) \le C\, n^{-s_* p/(2s_*+2m+3)} .$$

Remark 4.1. As usual in linear wavelet estimation, we distinguish in Theorem 4.1 two different zones: the homogeneous zone corresponding to $r \geq p$, and the inhomogeneous zone corresponding to $p > r$ (following the classification of [23, Remark 10.4]). For the homogeneous zone, we obtain the rate of convergence $u_{m,n} = n^{-sp/(2s+2m+3)}$ whereas for the inhomogeneous zone, $u_{m,n} = n^{-(s-1/r+1/p)p/(2(s-1/r+1/p)+2m+3)}$ which is slower than the previous one. Observe that these rates of convergence are similar to those attained by wavelet estimators for some inverse problems (see, e.g., [29], [24] and [12] for deconvolution problems).

Theorem 4.2 (L_p-risk for $\hat{f}^{(m)}_{hard}$)**.** *Consider* (1.1) *under (A1), (A2) and (A3). Let $\hat{f}^{(m)}_{hard}$ be* (3.5)*. Suppose that $f^{(m)} \in B^s_{r,q}(M)$ with $s > 0$, $r \geq 1$ and $q \geq 1$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\int_0^1 \left(\hat{f}^{(m)}_{hard}(x) - f^{(m)}(x)\right)^p dx\right) \leq C\,\varphi_{n,m}\ ,$$

where

$$\varphi_{n,m} = \begin{cases} \left(\dfrac{\ln n}{n}\right)^{sp/(2s+2m+3)}, & \text{for } rs > (m+3/2)(p-r), \\ \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p\,/\,(2s-2/r+2m+3)}, & \text{for } rs < (m+3/2)(p-r), \\ \left(\dfrac{\ln n}{n}\right)^{(s-1/r+1/p)p\,/\,(2s-2/r+2m+3)} (\ln n)^{(p-r/q)_+}, & \text{for } rs = (m+3/2)(p-r). \end{cases}$$

We see in Theorems 4.1 and 4.2 that

- over the homogeneous zone (i.e., $r \geq p$), $\hat{f}^{(m)}_{hard}$ attains a rate of convergence close to the one of $\hat{f}^{(m)}_{lin}$, i.e., $n^{-sp/(2s+2m+3)}$ (the only difference is a logarithmic term).
- over the inhomogeneous zone (i.e., $p > r$), $\hat{f}^{(m)}_{hard}$ attains a better rate of convergence than the one of $\hat{f}^{(m)}_{lin}$. From an asymptotic point of view, the difference is really significant.

Naturally, taking into account that $\hat{f}^{(m)}_{hard}$ is adaptive, it is preferable to $\hat{f}^{(m)}_{lin}$ in the estimation of $f^{(m)}$.

Remark 4.2. The optimal choice of the threshold κ is difficult to explicit because it depends on numerous constants including those in (A2) and (A3), some norms of the elements of the wavelet basis and the universal constants appearing in Bernstein inequality (see Proposition 6.2). The knowledge of these constants is however determinant for the knowledge of κ and, a fortiori, for the adaptivity of $\hat{f}^{(m)}_{hard}$.

Remark 4.3. Note that Theorem 4.2 taken with $p = 2$ and $m = 0$ coincides with [1, Theorem 4.2] taken with $w(x) = 1$.

Perspectives. A possible extension of this work will be to consider more complex thresholding technique as the block thresholding one (see, e.g., [6] and [13]). Moreover, to ensure that $n^{-sp/(2s+2m+3)}$ is the optimal in the minimax sense, lower bounds must be proved. However, important technical difficulties related to the estimation of $f^{(m)}$ (not only f) appear. All these aspects need further investigations that we leave for a future work.

5. SIMULATION STUDY

We investigate the performances of three wavelets estimators: the linear wavelet estimator (3.4) defined with $j_0 = 7$ (which is an arbitrary choice since s is unknown), the hard thresholding wavelet estimator 3.5 defined with the "universal threshold constant" $\kappa = \hat{\sigma}\sqrt{2}$, where $\hat{\sigma}$ is the standard deviation of the estimated wavelet coefficients (see [17]) and a linear wavelet estimator after local linear smoothing.

Remark 5.1. As noticed in [33], the smooth linear wavelet estimator is motivated by the fact that, when $f^{(m)}$ is smoother than the decomposing wavelet (or the sample size is small), the wavelet shrinkage estimators may contain abusive peaks and artifacts. A possible solution is to consider another smoothing method such as the local linear regression smoother introduced by [20, 21] which enjoys good sampling properties and high minimax efficiency. The construction of the considered estimator is based on [21, eq (2.1)-(2.4)], where Y_j is the wavelet linear estimator (3.4) with $j_0 = j$, $X_j = j/n$, K denotes the Gaussian kernel and $h = 0.08$. Note that we do not claim any theoretical properties of this estimator in this study.

The quality of the estimated density is measured by ANorm which are obtained by following formula

$$ANorm = \frac{1}{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{n}\left(\hat{f}_l^{(m)}(i/n) - f_l^{(m)}(i/n)\right)^2\right)^{1/2},$$

where N is the number of replications and $\hat{f}_l^{(m)}$ is estimator of $f_l^{(m)}$ in three state linear, hard threshold and smoothing methods. We select $N = 100$ and $m \in \{0, 1\}$ at (ANorm) formula. The codes were written in MATLAB software and use Daubechies-Lagarias algorithm for calculating various orthonormal wavelets.

In two examples, we consider samples from a Beta distribution and from a mixture of two Beta distributions.

In both of these examples, the smooth linear wavelet estimators is better than others. On the other hand, hard thresholding estimator (see (3.5)) works better than the linear estimator (see (3.4)).

Example 1. We generate samples $X_1, ..., X_n$ from a Beta distribution $Beta(\alpha, \beta)$ with parameters $\alpha = 3$ and $\beta = 3$ with size $n = 1000$. Also we generate $n = 1000$ samples from uniform distribution on $[0, 1]$ that are independent of the X_i's to produce multiplicative censoring. Then we estimate original density using various wavelet methods for derivatives of order $m \in \{0, 1\}$. Fig. 1 shows the original density and Fig. 2 its derivative.

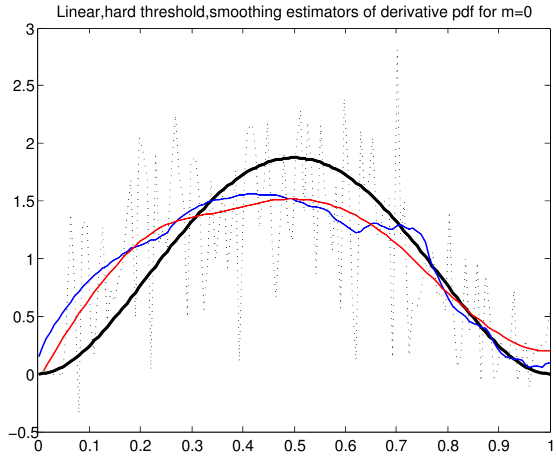

Figure 1: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Density estimation for Beta distribution).

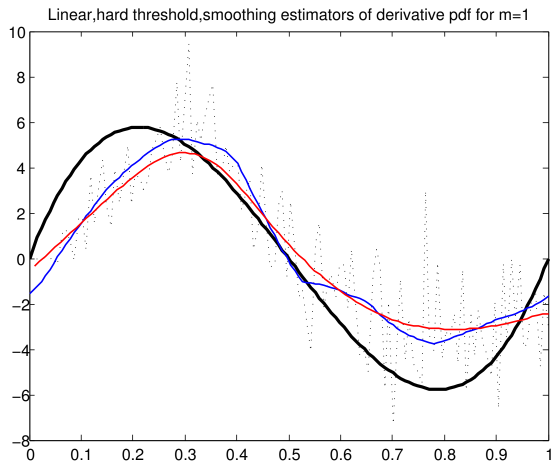

Figure 2: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Derivative estimation for Beta distribution).

The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line.

With obvious ANorm and standard deviation in Tables 1 and 2, we conclude when sample size increases, ANorm is smaller and we have better performance.

Table 1: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 0$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	12.7357 (0.9611)	9.2113 (0.7728)	6.8497 (0.5323)	5.2332 (0.4201)
Hard Thresholding	6.7080 (1.6041)	5.0373 (1.1540)	3.8189 (0.9575)	3.2546 (0.6949)
Smoothing	3.0102 (1.0213)	2.5472 (0.7237)	2.4273 (0.4855)	2.3779 (0.4057)

Table 2: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 1$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	40.4498 (3.3822)	31.5682 (2.5285)	25.8549 (1.6848)	22.2075 (1.0914)
Hard Thresholding	23.1572 (4.4096)	20.8959 (3.2689)	19.3508 (1.7315)	18.6700 (1.2586)
Smoothing	19.0204 (3.1333)	18.6047 (1.9025)	18.3701 (1.5284)	18.1202 (1.0132)

Example 2. In this example, we consider mixture Beta distribution. We generate $n = 1000$ samples $X_1, ..., X_n$ such that $f \sim (1/3)\, Beta(4,6) + (2/3)\, Beta(3,4)$ and proceed as the previous example. Fig. 3 and Fig. 4 show plot from defined estimators.

We calculated ANorm and standard deviation in Tables 3 and 4 for different values of n.

Table 3: Computed values for ANorm and (Standard deviation) with Beta mixture distribution for $m = 0$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	13.0786 (1.1155)	9.2829 (0.9413)	6.8973 (0.6469)	5.1470 (0.4214)
Hard Thresholding	9.0021 (1.8353)	6.0055 (1.4657)	4.5009 (0.9155)	3.7527 (0.5689)
Smoothing	3.0145 (1.0037)	2.7518 (0.8104)	2.7084 (0.6871)	2.5294 (0.3554)

Table 4: Computed values for ANorm and (Standard deviation) with Beta distribution for $m = 1$.

Estimation Methods	ANorm and (Standard deviation)			
	$n = 256$	$n = 512$	$n = 1024$	$n = 2048$
Linear	64.1098 (7.1914)	48.2251 (4.7379)	36.6836 (2.8821)	29.3822 (2.0521)
Hard Thresholding	32.9540 (10.2685)	26.2743 (7.7777)	25.0095 (4.1568)	21.9653 (2.2657)
Smoothing	20.5346 (4.5341)	20.2737 (4.0706)	19.6435 (2.3666)	19.3836 (1.9703)

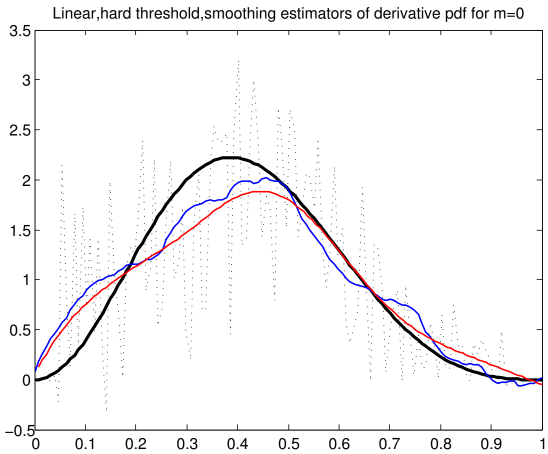

Figure 3: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Density estimation for Beta mixture distribution).

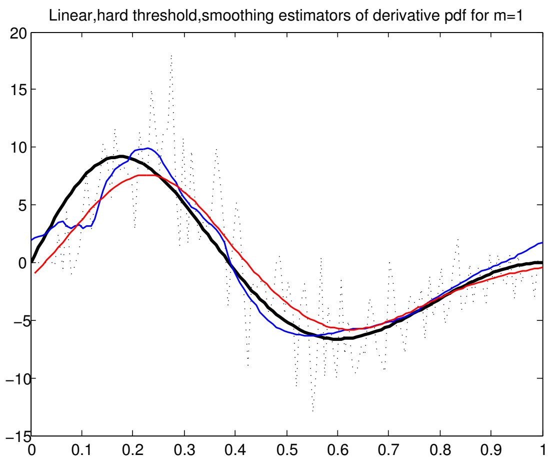

Figure 4: The original density with black line, linear estimator with dotted line, hard threshold estimator with blue line and smoothing estimator is red line. (Derivative estimation for Beta mixture distribution).

6. PROOFS

In this section, C denotes any constant that does not depend on j, k and n. Its value may change from one term to another and may depend on ϕ or ψ.

This section is organized as follows. Firstly, we introduce two auxiliary results on some properties of (3.2) and (3.3) at the heart of the proofs of our main theorems.

Proposition 6.1. *Let $p \geq 1$. For any integer $j \geq \tau$ such that $2^j \leq n$ and any $k \in \{0, ..., 2^j - 1\}$, let $\hat{c}_{j,k}^{(m)}$ be (3.2), $\hat{d}_{j,k}^{(m)}$ be (3.3), $c_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\phi_{j,k}(x)dx$ and $d_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\psi_{j,k}(x)dx$. Then, under (A1), (A2) and (A3), there exists a constant $C > 0$ such that*

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

and

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

Proposition 6.2. *Let $p \geq 1$. For any integer $j \geq \tau$ such that $2^j \leq n/\ln n$ and any $k \in \{0, ..., 2^j - 1\}$, let $\hat{d}_{j,k}^{(m)}$ be (3.3) and $d_{j,k}^{(m)} = \int_0^1 f^{(m)}(x)\psi_{j,k}(x)dx$. Then, under (A1), (A2) and (A3), there exists a constant $\kappa > 0$ such that*

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

Proof of Proposition 6.1: For convenience, let us prove the second inequality, the proof of the first one is identical.

For the sake of simplicity, for any $i \in \{1, ..., n\}$, set

$$Q_{i,j,k}^{(m)} = (-1)^m\left((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right)$$

and

$$U_i = Q_{i,j,k}^{(m)} - d_{j,k}^{(m)}\ .$$

Then we can write

$$\mathbb{E}\left((\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)})^{2p}\right) = \frac{1}{n^{2p}}\,\mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right). \tag{6.1}$$

Let us now investigate the bound of this expectation via the Rosenthal inequality presented below (see [34]).

Lemma 6.1 (Rosenthal's inequality)**.** *Let n be a positive integer, $\gamma \geq 2$ and $U_1, ..., U_n$ be n zero mean i.i.d. random variables such that $\mathbb{E}(|U_1|^\gamma) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left(\left|\sum_{i=1}^{n} U_i\right|^\gamma\right) \leq C \max\Big(n\,\mathbb{E}\big(|U_1|^\gamma\big),\, n^{\gamma/2}\left(\mathbb{E}(U_1^2)\right)^{\gamma/2}\Big)\,.$$

Observe that $U_1, ..., U_n$ are *i.i.d.* and, since $\mathbb{E}(Q_{i,j,k}^{(m)}) = d_{j,k}^{(m)}$, $\mathbb{E}(U_1) = 0$.

Since $Y_1(\Omega) = [0,1]$, we have

$$\begin{aligned} \left|Q_{i,j,k}^{(m)}\right| &\leq \left|(\psi_{j,k})^{(m)}(Y_i)\right| + \left|Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right| \\ &\leq \left|(\psi_{j,k})^{(m)}(Y_i)\right| + \left|(\psi_{j,k})^{(m+1)}(Y_i)\right|\,. \end{aligned} \tag{6.2}$$

Let $\upsilon \geq 1$. It follows from $\mathbb{E}(Q_{i,j,k}^{(m)}) = d_{j,k}^{(m)}$, the Hölder inequality and (6.2) that

$$\begin{aligned} \mathbb{E}\big(|U_1|^\upsilon\big) &\leq C\,\mathbb{E}\big(|Q_{1,j,k}^{(m)}|^\upsilon\big) \\ &\leq C\,\Big(\mathbb{E}\big(\left|(\psi_{j,k})^{(m)}(Y_1)\right|^\upsilon\big) + \mathbb{E}\big(\left|(\psi_{j,k})^{(m+1)}(Y_1)\right|^\upsilon\big)\Big)\,. \end{aligned} \tag{6.3}$$

Using (A3), $(\psi_{j,k})^{(m)}(x) = 2^{j(2m+1)/2}\psi^{(m)}(2^j x - k)$ and doing the change of variables $y = 2^j x - k$, we have

$$\begin{aligned} \mathbb{E}\big(\left|(\psi_{j,k})^{(m)}(Y_1)\right|^\upsilon\big) &= \int_0^1 \left|(\psi_{j,k})^{(m)}(x)\right|^\upsilon g(x)\,dx \;\leq C\int_0^1 \left|(\psi_{j,k})^{(m)}(x)\right|^\upsilon dx \\ &= C\,2^{j\upsilon(2m+1)/2}\int_0^1 \left|\psi^{(m)}(2^j x - k)\right|^\upsilon dx \\ &= C\,2^{j(\upsilon(2m+1)/2-1)}\int_{-k}^{2^j-k} \left|\psi^{(m)}(y)\right|^\upsilon dy \;\leq\; C\,2^{j(\upsilon(2m+1)/2-1)}\,. \end{aligned} \tag{6.4}$$

In a similar way, we prove that

$$\mathbb{E}\big(\left|(\psi_{j,k})^{(m+1)}(Y_1)\right|^\upsilon\big) \leq C\,2^{j(\upsilon(2m+3)/2-1)}\,. \tag{6.5}$$

Putting (6.3), (6.4) and (6.5) together, we obtain

$$\mathbb{E}\big(|U_1|^\upsilon\big) \leq C\,2^{j(\upsilon(2m+3)/2-1)}\,. \tag{6.6}$$

Using the Rosenthal inequality with $U_1, ..., U_n$, $\gamma = 2p$ and $2^j \leq n$, we have

$$\begin{aligned} \mathbb{E}\left(\left(\sum_{i=1}^{n} U_i\right)^{2p}\right) &\leq C\,\max\Big(n\,\mathbb{E}(U_1^{2p}),\, n^p\left(\mathbb{E}(U_1^2)\right)^p\Big) \\ &\leq C\,\max\Big(n\,2^{j((2m+3)p-1)},\, n^p\,2^{j(2m+2)p}\Big) \\ &\leq C\,n^p\,2^{j(2m+2)p}\,. \end{aligned} \tag{6.7}$$

By (6.1) and (6.7), we have

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

Similarly, we prove that

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

The proof of Proposition 6.1 is complete. □

Proof of Proposition 6.2: For the sake of simplicity, for any $i \in \{1, ..., n\}$, set

$$Q_{i,j,k}^{(m)} = (-1)^m \left((\psi_{j,k})^{(m)}(Y_i) + Y_i(\psi_{j,k})^{(m+1)}(Y_i)\right)$$

and

$$U_i = Q_{i,j,k}^{(m)} - d_{j,k}^{(m)}\,.$$

Then, for any $\kappa > 0$, we can write

(6.8)
$$\mathbb{P}\left(|\hat{d}_{j,k}^{(m)} - d_{j,k}^{(m)}| \ge \frac{\kappa}{2}\,2^{j(m+1)}\sqrt{\frac{\ln n}{n}}\right) = \mathbb{P}\left(\left|\sum_{i=1}^n U_i\right| \ge C\,\frac{\kappa}{2}\,2^{j(m+1)}\sqrt{n\ln n}\right).$$

Let us now explore the bound of this probability via the Bernstein inequality described below (see [30]).

Lemma 6.2 (Bernstein's inequality)**.** *Let n be a positive integer and $U_1, ..., U_n$ be n i.i.d. zero mean independent random variables such that there exists a constant $M > 0$ satisfying $|U_1| \le M < \infty$. Then, for any $\upsilon > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n U_i\right| \ge \upsilon\right) \le 2\exp\left(-\frac{\upsilon^2}{2\left(n\,\mathbb{E}(U_1^2) + \upsilon M/3\right)}\right).$$

Observe that $U_1, ..., U_n$ are *i.i.d.* and, since $\mathbb{E}(Q_{i,j,k}^{(m)}) = d_{j,k}^{(m)}$, $\mathbb{E}(U_1) = 0$.

Since $Y_1(\Omega) = [0,1]$, $(\psi_{j,k})^{(m)}(x) = 2^{j(2m+1)/2}\psi^{(m)}(2^j x - k)$, $\sup_{y\in[0,1]} |(\psi_{j,k})^{(m)}(y)| \le C2^{j(2m+1)/2}$ and $\sup_{y\in[0,1]} |(\psi_{j,k})^{(m+1)}(y)| \le C2^{j(2m+3)/2}$, we have

$$\begin{aligned}|Q_{1,j,k}^{(m)}| &\le \left|(\psi_{j,k})^{(m)}(Y_1)\right| + \left|Y_1(\psi_{j,k})^{(m+1)}(Y_1)\right| \\ &\le C\left(\sup_{y\in[0,1]}\left|(\psi_{j,k})^{(m)}(y)\right| + \sup_{y\in[0,1]}\left|(\psi_{j,k})^{(m+1)}(y)\right|\right) \le C\,2^{j(2m+3)/2}\,.\end{aligned}$$

Observe that, thanks to (A2) and the Cauchy–Schwarz inequality,

$$|d_{j,k}^{(m)}| \le \left(\int_0^1 \left(f^{(m)}(x)\right)^2 dx\right)^{1/2}\left(\int_0^1 \left(\psi_{j,k}(x)\right)^2 dx\right)^{1/2} \le C\,.$$

Using $2^j \le n/\ln n$, we have

$$\begin{aligned}|U_1| \;&\le\; C\big(\big|Q^{(m)}_{1,j,k}\big| + \big|d^{(m)}_{j,k}\big|\big) \le C\big(2^{j(2m+3)/2} + C\big)\\ &=\; C\,2^{j(2m+3)/2} \le C\,2^{j(m+1)}\sqrt{\frac{n}{\ln n}}\;.\end{aligned}$$

It follows from (6.6) that

$$\mathbb{E}(U_1^2) \;\le\; C\,2^{j(2m+2)}\;.$$

The Bernstein inequality applied with $U_1,...,U_n$ and $\upsilon = (\kappa/2)\,2^{j(m+1)}\sqrt{n\ln n}$ gives

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^n U_i\right| \ge \upsilon\right) &\le 2\,\exp\left(-\frac{\upsilon^2}{2\big(n\,\mathbb{E}(U_1^2) + \upsilon M/3\big)}\right)\\ &\le\; 2\,\exp\left(-\frac{(\kappa/2)^2\,2^{j(2m+2)}\,n\,\ln n}{C n 2^{j(2m+2)} + C(\kappa/2)\,2^{j(m+1)}\sqrt{n\ln n}\;2^{j(m+1)}\sqrt{n/\ln n}}\right) \qquad (6.9)\\ &=\; 2\,n^{-C\frac{\kappa^2}{1+\kappa}}\;.\end{aligned}$$

By (6.8) and (6.9), there exists a constant $\kappa > 0$ such that

$$\mathbb{P}\left(\big|\hat d^{(m)}_{j,k} - d^{(m)}_{j,k}\big| \ge \frac{\kappa}{2}\,2^{j(m+1)}\sqrt{\frac{\ln n}{n}}\right) \;\le\; 2\,n^{-C\frac{\kappa^2}{1+\kappa}} \;\le\; 2\left(\frac{\ln n}{n}\right)^p.$$

Proposition 6.2 is proved. □

Proof of Theorem 4.1: We expand the function $f^{(m)}$ on $\mathcal{S}$ as

$$f^{(m)}(x) \;=\; \sum_{k=0}^{2^{j_0}-1} c^{(m)}_{j_0,k}\,\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d^{(m)}_{j,k}\,\psi_{j,k}(x)\;,$$

where $c^{(m)}_{j_0,k} = \int_0^1 f^{(m)}(x)\,\phi_{j_0,k}(x)\,dx$ and $d^{(m)}_{j,k} = \int_0^1 f^{(m)}(x)\,\psi_{j,k}(x)\,dx$.

We have

$$\mathbb{E}\left(\int_0^1\Big(\hat f^{(m)}_{lin}(x) - f^{(m)}(x)\Big)^p\,dx\right) \le 2^{p-1}(A+B)\;, \qquad (6.10)$$

where

$$A \;=\; \mathbb{E}\left(\int_0^1\left(\sum_{k=0}^{2^{j_0}-1}\big(\hat c^{(m)}_{j_0,k} - c^{(m)}_{j_0,k}\big)\,\phi_{j_0,k}(x)\right)^p dx\right)$$

and

$$B \;=\; \int_0^1\left(\sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d^{(m)}_{j,k}\,\psi_{j,k}(x)\right)^p dx\;.$$

Let us now introduce a L_p-norm result for wavelets.

Lemma 6.3. *Let $p \geq 1$. For any sequence of real number $(\theta_{j,k})_{j,k}$, there exists a constant $C > 0$ such that*

$$\int_0^1 \left(\sum_{k=0}^{2^j-1} \theta_{j,k}\, \phi_{j,k}(x) \right)^p dx \;\leq\; C\, 2^{j(p/2-1)} \sum_{k=0}^{2^j-1} |\theta_{j,k}|^p .$$

The proof can be found in, e.g., [23, Proposition 8.3].

Lemma 6.3, Proposition 6.1 and the Cauchy–Schwarz inequality yield

$$\begin{aligned} A \;&\leq\; C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E}\left(\left(\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k}\right)^p \right) \\ &\leq\; C\, 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \left(\mathbb{E}\left(\left(\hat{c}^{(m)}_{j_0,k} - c^{(m)}_{j_0,k}\right)^{2p} \right) \right)^{1/2} \qquad (6.11) \\ &\leq\; C\, 2^{j_0(p/2-1)}\, 2^{j_0}\, 2^{j_0(m+1)p} \frac{1}{n^{p/2}} \;=\; C \left(\frac{2^{j_0(2m+3)}}{n} \right)^{p/2} . \end{aligned}$$

On the other hand, using $f^{(m)} \in B^s_{r,q}(M)$ and proceeding as in [18, eq (24)], we have

$$B \;\leq\; C\, 2^{-j_0 s_* p} . \qquad (6.12)$$

It follows from (6.10), (6.11), (6.12) and the definition of j_0 that

$$\begin{aligned} \mathbb{E}\left(\int_0^1 \left(\hat{f}^{(m)}_{lin}(x) - f^{(m)}(x) \right)^p dx \right) \;&\leq\; C \left(\left(\frac{2^{j_0(2m+3)}}{n} \right)^{p/2} + 2^{-j_0 s_* p} \right) \\ &\leq\; C\, n^{-s_* p/(2s_*+2m+3)} . \end{aligned}$$

This ends the proof of Theorem 4.1. □

Proof of Theorem 4.2: Theorem 4.2 is a consequence of Theorem 6.1 below by taking with $\nu = m+1$ and using Propositions 6.1 and 6.2 above. □

Theorem 6.1. *Let $h \in L_2([0,1])$ be an unknown function to be estimated from n observations and (2.1) its wavelet decomposition. Let $\hat{c}_{j,k}$ and $\hat{d}_{j,k}$ be estimators of $c_{j,k}$ and $d_{j,k}$ respectively such that there exist three constants $\nu > 0$, $C > 0$ and $\kappa > 0$ satisfying*

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

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

and

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

Concentration inequality: for any $j\ge\tau$ *such that* $2^j\le n/\ln n$ *and* $k\in\{0,...,2^j-1\}$,

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

Let us define the hard thresholding wavelet estimator of h *by*

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

where j_1 *is the integer satisfying* $(n/\ln n)^{1/(2\nu+1)}<2^{j_1+1}\le 2\,(n/\ln n)^{1/(2\nu+1)}$.

Suppose that $h\in B^s_{r,q}(M)$ *with* $s>0$, $r\ge1$ *and* $q\ge1$. *Then there exists a constant* $C>0$ *such that*

$$\mathbb{E}\left(\int_0^1\left(\hat{h}(x)-h(x)\right)^{p}dx\right)\le C\,\Theta_{n,\nu}\ ,$$

where

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

Theorem 6.1 does not appear in this form in the literature but can be proved using similar arguments to [25, Theorem 5.1] for a bound of the L_p-risk and [12, Theorem 4.2] for the determination of the rates of convergence.

ACKNOWLEDGMENTS

The authors thank two anonymous referees for their thorough and useful comments.

REFERENCES

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

[2] Andersen, K. and Hansen, M. (2001). Multiplicative censoring: density estimation by a series expansion approach, *Journal of Statistical Planning and Inference*, **98**, 137–155.

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

[4] Bhattacharya, P.K. (1967). Estimation of a probability density function and its derivatives, *Sankhya, Serie A*, **29**, 373–382.

[5] Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model, *Bernoulli*, **11**(2), 309–340.

[6] Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898–924.

[7] Chabert, M.; Ruiz, D. and Tourneret, J.-Y. (2004). Optimal wavelet for abrupt change detection in multiplicative noise. In "IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004", Proceedings, (ICASSP '04), vol. 2, pp. ii 1089-92. IEEE, May 2004.

[8] Chacón, J.E.; Duong, T. and Wand, M.P. (2011). Asymptotics for general multivariate kernel density derivative estimators, *Statistica Sinica*, to appear.

[9] Chaubey, Y.P.; Doosti, H. and Prakasa Rao, B.L.S. (2006). Wavelet based estimation of the derivatives of a density with associated variables, *International Journal of Pure and Applied Mathematics*, **27**(1), 97–106.

[10] Chaubey, Y.P.; Doosti, H. and Prakasa Rao, B.L.S. (2008). Wavelet based estimation of the derivatives of a density for a negatively associated process, *Journal of Statistical Theory and Practice*, **2**(3), 453–463.

[11] Chaubey, Y.P.; Chesneau, C. and Doosti, H. (2011). On linear wavelet density estimation: some recent developments, *Journal of the Indian Society of Agricultural Statistics*, **65**(2), 169–179.

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

[13] Chesneau, C.; Fadili, M.J. and Starck, J.-L. (2010). Stein Block Thresholding For Image Denoising, *Applied and Computational Harmonic Analysis*, **28**(1), 67–88.

[14] 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.

[15] Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems, *Statistica Sinica*, **16**, 773–788.

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

[17] DONOHO, D.L. and JOHNSTONE, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrica*, **81**, 425–455.

[18] DONOHO, D.L.; JOHNSTONE, I.M.; KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding, *The Annals of Statistics*, **24**, 508–539.

[19] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problem, *Ann. Statist.*, **19**, 1257–1272.

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

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

[22] FAN, J. and LIU, Y. (1997). A note on asymptotic normality for deconvolution kernel density estimators, *Sankhya*, **59**, 138–141.

[23] HÄRDLE, W.; KERKYACHARIAN, G.; PICARD, D. and TSYBAKOV, A. (1998). *Wavelet, Approximation and Statistical Applications*, Lectures Notes in Statistics, 129, New York, Springer Verlag.

[24] JOHNSTONE, I.; KERKYACHARIAN, G.; PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting, *Journal of the Royal Statistical Society, Serie B*, **66**(3), 547–573.

[25] KERKYACHARIAN, G. and PICARD, D. (2000). Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors), *Test*, **9**(2), 283–345.

[26] LACOUR, C. (2006). Rates of convergence for nonparametric deconvolution, *C. R. Acad. Sci. Paris Ser. I Math.*, **342**(11), 877–882.

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

[28] MEYER, Y. (1992). *Wavelets and Operators*, Cambridge University Press, Cambridge.

[29] PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution, *The Annals of Statistics*, **27**, 2033–2053.

[30] PETROV, V.V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*, Oxford: Clarendon Press.

[31] PRAKASA RAO, B.L.S. (1996). Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bull. Inform. Cyb.*, **28**, 91–100.

[32] PRAKASA RAO, B.L.S. (2010). Wavelet linear estimation for derivatives of a density from observations of mixtures with varying mixing proportions, *Indian J. Pure Appl. Math.*, **41**(1), 275–291.

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

[34] 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.

[35] SCHUSTER, E.F. (1969). Estimation of a probability function and its derivatives, *Ann. Math. Statist.*, **40**, 1187–1195.

[36] Singh, R.S. (1977). Applications of estimators of a density and its derivatives to certain statistical problems, *J. Roy. Statist. Soc. B*, **39**, 357–363.

[37] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation, *Biometrika*, **76**, 751–761.

[38] Vardi, Y. and Zhang, C.H. (1992). Large sample study of empirical distributions in a random multiplicative censoring model, *The Annals of Statistics*, **20**, 1022–1039.

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