



HAL
open science

On a Plug-In Wavelet Estimator for Convolutions of Densities

Christophe Chesneau, Fabien Navarro

► **To cite this version:**

Christophe Chesneau, Fabien Navarro. On a Plug-In Wavelet Estimator for Convolutions of Densities. *Journal of Statistical Theory and Practice*, 2014, 8 (4), 10.1080/15598608.2013.824824. hal-00625912v2

HAL Id: hal-00625912

<https://hal.science/hal-00625912v2>

Submitted on 4 Apr 2013

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.

ON A PLUG-IN WAVELET ESTIMATOR FOR CONVOLUTIONS OF DENSITIES

CHRISTOPHE CHESNEAU^a AND FABIEN NAVARRO^{a,b}

^a*LMNO CNRS-Université de Caen, Campus II, 14032 Caen Cedex, France*

^b*GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen Cedex, France*

April 4, 2013

Abstract

The nonparametric estimation of the m -fold convolution power of an unknown function f is considered. We introduce an estimator based on a plug-in approach and a wavelet hard thresholding estimator. We explore its theoretical asymptotic performances via the mean integrated squared error assuming that f has a certain degree of smoothness. Applications and numerical examples are given for the standard density estimation problem and the deconvolution density estimation problem.

Key words and phrases: density estimation; convolutions; plug-in estimator; wavelets; hard thresholding; rates of convergence

1 Introduction

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, f be an unknown function related to n *i.i.d.* random variables Z_1, \dots, Z_n and $m \geq 2$ be a fixed integer. We aim to estimate the m -fold convolution power of f

$$g(x) = \star^m f(x) = \int \dots \int f(x - u_2 - \dots - u_m) f(u_2) \dots f(u_m) du_2 \dots du_m \quad (1.1)$$

from Z_1, \dots, Z_n . For the case where $m = 2$, note that $\star^2 f(x) = (f \star f)(x) = \int f(x - t) f(t) dt$.

Probably the most famous example concerns the density estimation problem where f is the density of Z_1 and g is the density of $S = \sum_{i=1}^m Z_i$. Many quantities of interest in actuarial or financial sciences involve sums of random variables. For example, in the individual risk model, the total amount of claims on a portfolio of insurance contracts is modelled as the sum of all claims on the individual policies. Therefore, probability density functions of sums of random variables are of particular interest. A typical example is the sum of insurance claims, where $(Z_i)_{i=1, \dots, n}$ are individual insurance claims and $S = \sum_{i=1}^m Z_i$ is the sum of m claims and m could be interpreted as the expected number of claims in a specified period (e.g. one month). As an example, we refer to [Frees \(1994\)](#) which studied the total charges for female patients admitted to the Wisconsin Hospital for circulatory disorders during a year. Another detailed application in the field of health insurance can be found in [Panjer and Willmot \(1992\)](#). Methods and results can be found in [Frees \(1994\)](#), [Saavedra and Cao \(2000\)](#), [Ahmad and Fan \(2001\)](#), [Ahmad and Mugdadi \(2003\)](#), [Prakasa Rao \(2004\)](#), [Schick and Wefelmeyer \(2004, 2007\)](#), [Du and Schick \(2007\)](#) and [Giné and Mason \(2007\)](#). In particular, [Saavedra and Cao \(2000\)](#) have introduced the natural plug-in estimator $\hat{g} = \star^m \hat{f}$, where \hat{f} denotes a kernel estimator.

In this study, considering the general form of the problem, we propose to extend the approach of [Saavedra and Cao \(2000\)](#) to the wavelet hard thresholding estimators. We are interested on such estimators because they achieve a high degree of adaptivity and capability of handling the singularities of the unknown function. We refer to e.g. [Antoniadis \(1997\)](#), [Härdle et al. \(1998\)](#) and [Vidakovic \(1999\)](#) for the details and discussions on their advantages over traditional methods. We study the asymptotic performance of our estimator by considering the mean integrated squared error (MISE) and assuming that f belongs to a wide class of unknown functions (the Besov balls). The obtained rate of convergence is of

the form $(\ln n/n)^\omega$, where $\omega \in (0, 1)$ depends on the regularity of f and the considered statistical model. Let us mention that it corresponds to the standard one attained by the hard thresholding estimator for f from various statistical models. Further details can be found in [Kerkycharian and Picard \(2000\)](#).

If we focus our attention on the density estimation model, the obtained rate of convergence is a bit slower to the one reached by the kernel estimator of [Saavedra and Cao \(2000\)](#) (i.e. the parametric rate $1/n$). However, our result has the advantage to be applied for a wide class of densities. We do not need to have f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$, f'' and $f^{(4)}$ are integrable as in ([Saavedra and Cao, 2000](#), Theorem 3); the assumption that f belongs to Besov balls provides more flexibility.

Finally we illustrate our theoretical result by considering two different density estimation problems: the standard density one (described above) and the deconvolution density one. To the best of our knowledge, the consideration of the deconvolution density model in this setting is new but in many applications, errors-in-variables could arise, for example, measurement error is a problem inherent in health insurance claims data. Thus, the deconvolution density estimation problem can be viewed as a natural extension. A comprehensive simulation study supports our theoretical findings. In particular, we show that the practical performance of our wavelet estimator compares favorably to the standard ones (i.e. those of [Frees \(1994\)](#) and [Saavedra and Cao \(2000\)](#)) for a wide variety of functions.

The paper is organized as follows. Our wavelet hard thresholding methodology and the Besov balls are presented in Section 2. Section 3 explores the MISE properties of our estimator. Applications of our theory and simulation results are presented in Section 4 for the standard density estimation and Section 5 for the deconvolution density estimation. Technical proofs are given in Section 6.

2 Wavelet estimators

First of all, we introduce some basics on wavelets. Then we develop our wavelet hard thresholding estimator and define the Besov balls.

2.1 Basics on wavelets

Let $N \geq 1$ be an integer, and ϕ and ψ be the initial wavelet functions of the Daubechies wavelets $db2N$. These functions have the particularity to be compactly supported and C^ν where ν is an integer depending on N .

From these wavelet two functions, we define $\phi_{j,k}$ and $\psi_{j,k}$ by

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

Then there exists an integer τ and a set of consecutive integers Λ_j with a length proportional to 2^j such that, for any integer $\ell \geq \tau$, the collection

$$\mathcal{B} = \{\phi_{\ell,k}, k \in \Lambda_\ell; \psi_{j,k}; j \in \mathbb{N} - \{0, \dots, \ell - 1\}, k \in \Lambda_j\},$$

is an orthonormal basis of $\mathbb{L}^2([-T, T]) = \{h : [-T, T] \rightarrow \mathbb{R}; \int_{-T}^T h^2(x)dx < \infty\}$.

Suppose that $h \in \mathbb{L}^2([-T, T])$. Then, for any integer $\ell \geq \tau$, we can write a formal expansion

$$h(x) = \sum_{k \in \Lambda_\ell} \alpha_{\ell,k} \phi_{\ell,k}(x) + \sum_{j=\ell}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x), \quad (2.1)$$

where $\alpha_{j,k}$ and $\beta_{j,k}$ are the wavelet coefficients of h defined by the integrals:

$$\alpha_{j,k} = \int h(x) \phi_{j,k}(x) dx, \quad \beta_{j,k} = \int h(x) \psi_{j,k}(x) dx. \quad (2.2)$$

For details about wavelet basis, we refer to [Cohen et al. \(1993\)](#) and [Mallat \(2009\)](#).

2.2 Estimators

Let us consider the general estimation problem described in Section 1 and suppose that $f \in \mathbb{L}^2([-T, T])$. We expand the unknown function f on \mathcal{B} as (2.1). Let $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ be estimators of the wavelet coefficients $\alpha_{j,k} = \int f(x) \phi_{j,k}(x) dx$ and $\beta_{j,k} = \int f(x) \psi_{j,k}(x) dx$ respectively.

Following the general approach of [Kerkycharian and Picard \(2000\)](#), we suppose that there exist three constants $C > 0$, $\kappa > 0$ and $\delta > 0$ such that $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ satisfy, for any $j \in \{\tau, \dots, j_1\}$,

(i)

$$\mathbb{E}(|\hat{\alpha}_{j,k} - \alpha_{j,k}|^v) \leq C 2^{v\delta j} \left(\frac{\ln n}{n}\right)^{v/2}, \quad (2.3)$$

(ii)

$$\mathbb{E}(|\hat{\beta}_{j,k} - \beta_{j,k}|^v) \leq C 2^{v\delta j} \left(\frac{\ln n}{n}\right)^{v/2} \quad (2.4)$$

(iii)

$$\mathbb{P}\left(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \frac{\kappa}{2} 2^{\delta j} \sqrt{\frac{\ln n}{n}}\right) \leq C \left(\frac{\ln n}{n}\right)^4, \quad (2.5)$$

where $v = 4m - 4 \geq 4$ and j_1 is the integer satisfying

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

For a wide variety of models, one may construct $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ satisfying (2.3), (2.4) and (2.5). The parameter δ plays a major role for some inverse problems; in the standard statistical model (density, ...), it is often equal to 0. Examples are given in Sections 4 and 5.

Then we define the hard thresholding estimator \hat{f} by

$$\hat{f}(x) = \sum_{k \in \Lambda_\tau} \hat{\alpha}_{\tau,k} \phi_{\tau,k}(x) + \sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} \hat{\beta}_{j,k} \mathbf{1}_{\{|\hat{\beta}_{j,k}| \geq \kappa 2^{\delta j} \sqrt{\ln n/n}\}} \psi_{j,k}(x), \quad (2.6)$$

where, for any random event \mathcal{A} , $\mathbf{1}_{\mathcal{A}}$ is the indicator function on \mathcal{A} .

The idea of the hard thresholding rule in (2.6) is to make a term-by-term selection: only the "large" unknown wavelet coefficients of f which contain its main characteristics are estimated. Details can be found in e.g. Antoniadis (1997), Härdle et al. (1998) and Vidakovic (1999).

Using (2.6), we consider the following plug-in estimator for $g = \star^m f$:

$$\hat{g}(x) = \star^m \hat{f}(x). \quad (2.7)$$

To study its asymptotic performance, we need some smoothness assumptions on f . In this study, as usual in wavelet estimation, we suppose that f belongs to Besov balls defined below.

2.3 Besov balls

We say that $h \in B_{p,r}^s(M)$ with $M > 0$, $s > 0$, $p \geq 1$ and $r \geq 1$ if and only if there exists a constant $M^* > 0$ (depending on M) such that (2.2) satisfy

$$\left(\sum_{j=\tau}^{\infty} \left(2^{j(s+1/2-1/p)} \left(\sum_{k \in \Lambda_j} |\beta_{j,k}|^p \right)^{1/p} \right)^r \right)^{1/r} \leq M^*.$$

In this expression, s is a smoothness parameter and p and r are norm parameters. We consider such Besov balls essentially because of their executional expressive power. In particular, they contain the Hölder and Sobolev balls. See e.g. Meyer (1992) and (Härdle et al., 1998, Chapter 9).

3 Upper bound

Theorem 3.1 below investigates the rates of convergence for \hat{g} (2.7) under the MISE over Besov balls.

Theorem 3.1. *Consider the estimation problem and notations of Section 1. Suppose that $\text{supp } f \subseteq [-T, T]$, where $T > 0$ is a fixed constant, and there exists a constant $C > 0$ such that $\int |f(x)|^{4m-4} dx \leq C$. Let \hat{g} be (2.7) (under (2.3), (2.4) and (2.5)). Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 4 \text{ and } s > 0\}$ or $\{p \in [1, 4) \text{ and } s > \max((2\delta+1)/p, (4/p-1)(\delta+1/2))\}$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}.$$

The proof of Theorem 3.1 uses a suitable decomposition of the MISE and a result on the rates of convergence of \hat{f} under the \mathbb{L}_p -risk with $p \in \{2, 4\}$ over Besov balls derived from Kerkyacharian and Picard (2000). Let us mention that the control of both of the \mathbb{L}_2 -risk (MISE) and \mathbb{L}_4 -risk motivates the consideration of the wavelet hard thresholding estimator in the definition of \hat{g} (2.7) instead of other wavelet estimators (as the block thresholding ones introduced by Cai (1999)).

Theorem 3.1 shows that under mild assumptions on

- the model: only (2.3), (2.4) and (2.5) are required,
- f : only $\text{supp } f \subseteq [-T, T]$, $\int |f(x)|^{4m-4} dx \leq C$, $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 4 \text{ and } s > 0\}$ or $\{p \in [1, 4) \text{ and } s > \max((2\delta + 1)/p, (4/p - 1)(\delta + 1/2))\}$ (including the inhomogeneous zone of the Besov balls corresponding to $p \in [1, 2)$) are required,

the estimator \hat{g} attains a "fast" rate of convergence. "Fast" in the sense that it is close to the parametric rate $1/n$. However, we do not claim it to be optimal in the minimax sense. This point will be discussed for two particular density models in the next section.

It is important to mention that the rate of convergence $(\ln n/n)^{2s/(2s+2\delta+1)}$ is the near optimal one in the minimax sense for \hat{f} (not \hat{g}) under the MISE over Besov balls for various standard nonparametric setting (density model, nonparametric regression model, deconvolution density model, ...). See e.g. Härdle et al. (1998), Fan and Koo (2002) and Tsybakov (2004).

4 Application I: the density model

4.1 Upper bound

We observe n *i.i.d.* random variables Z_1, \dots, Z_n with common unknown density f . For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^m Z_i$ and g be the density of S . The goal is to estimate g from Z_1, \dots, Z_n . As mentioned in Section 1, such a problem has already been considered with kernel-type estimators and various settings by e.g. Frees (1994), Saavedra and Cao (2000), Ahmad and Fan (2001), Ahmad and Mugdadi (2003), Schick and Wefelmeyer (2004, 2007) and Du and Schick (2007).

Proposition 4.1 below investigates the rates of convergence of \hat{g} (2.7) constructed from a specific wavelet hard thresholding estimator \hat{f} under the MISE over Besov balls.

Proposition 4.1. *Consider the standard density model and the associated notations. Suppose that $\text{supp } f \subseteq [-T, T]$, where $T > 0$ is a fixed constant, and there exists a constant $C > 0$ such that $\sup_{x \in \mathbb{R}} f(x) \leq C$.*

Let \hat{g} be (2.7) with $\delta = 0$,

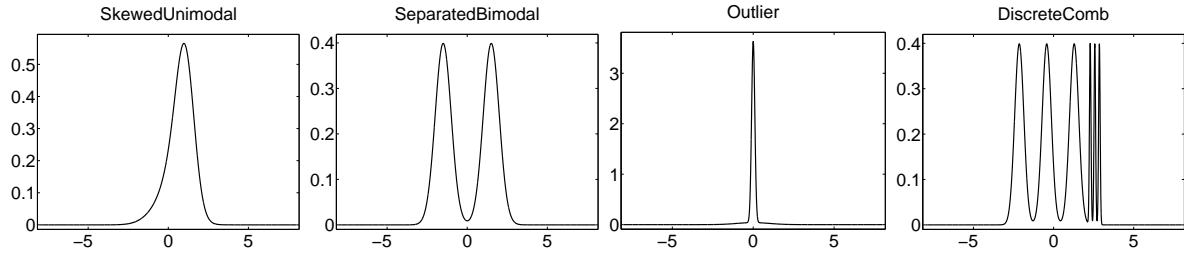
$$\hat{\alpha}_{j,k} = \frac{1}{n} \sum_{i=1}^n \phi_{j,k}(Z_i), \quad \hat{\beta}_{j,k} = \frac{1}{n} \sum_{i=1}^n \psi_{j,k}(Z_i). \quad (4.1)$$

Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 4 \text{ and } s > 0\}$ or $\{p \in [1, 4) \text{ and } s > \max(1/p, (2/p - 1/2))\}$. Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+1)}.$$

As noted in Section 1, the rate of convergence $(\ln n/n)^{2s/(2s+1)}$ is a bit slower than the one reached by the kernel estimator of Saavedra and Cao (2000) (i.e. the parametric rate $1/n$). The more s is large, the more they are close. The main contribution of Proposition 4.1 concerns the assumptions on f : we do not need to have f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$, f'' and $f^{(4)}$ are integrable as in (Saavedra and Cao, 2000, Theorem 3); the assumption that f belongs to Besov balls includes a wide class of functions which does not satisfy such assumptions.

Remark that $(\ln n/n)^{2s/(2s+1)}$ is the "near optimal" rates of convergence in the minimax sense for the standard density estimation problem for \hat{f} under the MISE over Besov balls. See (Donoho et al., 1996, Theorems 2 and 3).

Figure 1: Theoretical densities from $n = 1000$ samples Z_1, \dots, Z_n .

4.2 Simulation results

In the following simulation study, we have analyzed the performances of our adaptive wavelet estimation procedure on a family of normal mixture densities (“SeparatedBimodal”, “Outlier” and “DiscreteComb”, initially introduced in [Marron and Wand \(1992\)](#)) representing different degrees of smoothness (see Fig. 1). We have adapted the formulae given by [Marron and Wand \(1992\)](#) to simulate such densities by choosing the parameters so that

$$\min_l(q\mu_l - 3q\sigma_l) = -3q, \quad \max_l(q\mu_l + 3q\sigma_l) = 3q$$

where $l = 1, \dots, p$ with p the number of densities in the mixture and q a scaling parameter. Thereby, it is very unlikely to have values outside the interval $[-4q, 4q]$ ($\subseteq [-T, T]$) and we lose little by assuming compact support (see Fig. 1).

Since our estimation method is adaptive, we have chosen a predetermined threshold κ (universal thresholding, see e.g. [Donoho et al. \(1996\)](#)) for all the tests and the Symmlet wavelet with 6 vanishing moments was used throughout all experiments. The finest resolution level j_1 used in all our simulations was chosen to be the maximum resolution level allowed by the discretization. For each density, $n = 1000$ independent samples were generated and the MISE was approximated as an average of the Integrated Squared Error (ISE) over 100 replications. The m -fold convolution product of f with itself defined by (2.7) can be efficiently computed numerically using the fast Fourier transform, thanks to the convolution theorem. Typical reconstructions from a single simulation are depicted in Fig. 2 for $m = 1$, $m = 2$ and $m = 3$ respectively. One can see that our adaptive hard thresholding estimator is very effective to estimate each of the nine densities.

Then, we have compared the performance of our adaptive wavelet estimator to those of two different kernel-based estimators. The first one, presented in [Saavedra and Cao \(2000\)](#), is based on convolving kernel density estimators: $\hat{g} = \star^m \hat{f}$, where \hat{f} denotes a kernel estimator. The other one, introduced by [Frees \(1994\)](#), is the Frees type local U-statistic estimator defined as follow

$$\hat{g}(x) = \frac{1}{\binom{n}{m}b} \sum_{(n,m)} K\left(\frac{x - h(Z_{i_1}, \dots, Z_{i_m})}{b}\right), \quad (4.2)$$

where b is the bandwidth or smoothing parameter, K is a kernel function and $\sum_{(n,m)}$ denotes summation over all $\binom{n}{m}$ subsets. Recall that we have focused here on the interesting case where $h(Z_1, \dots, Z_m) = \sum_{i=1}^m Z_i$ (see [Frees \(1994\)](#) for applications).

In the sequel, we name the estimator of [Saavedra and Cao \(2000\)](#) by ‘Kernel’, the one of [Frees \(1994\)](#) by ‘Frees’ and our estimator by ‘Wavelet’.

In the case of *i.i.d.* random variables, the choice of the kernel is not crucial for density estimation. However, it is well known that the choice of the bandwidth is very important. Many procedures of bandwidth selection for density estimation have been developed in the literature (details can be found in [Mugdadi and Ahmad \(2004\)](#) where several methods are compared). Here, for both kernel-based estimators, we have been focused on a global bandwidth selector: the *rule of thumb* (ROT) bandwidth selector (see e.g. [Silverman \(1986\)](#)). [Ahmad and Fan \(2001\)](#) derived the asymptotic mean integrated square error and the optimal bandwidth for the Frees estimator (4.2). Thanks to ([Ahmad and Fan, 2001](#), Theorem 2.2) with $r = 2$ and the Gaussian kernel, the optimal bandwidth is given by

$$b_{ROT} = 1.06 \min(\hat{\sigma}, Q/1.34) \binom{n}{m}^{-1/5} \quad (4.3)$$

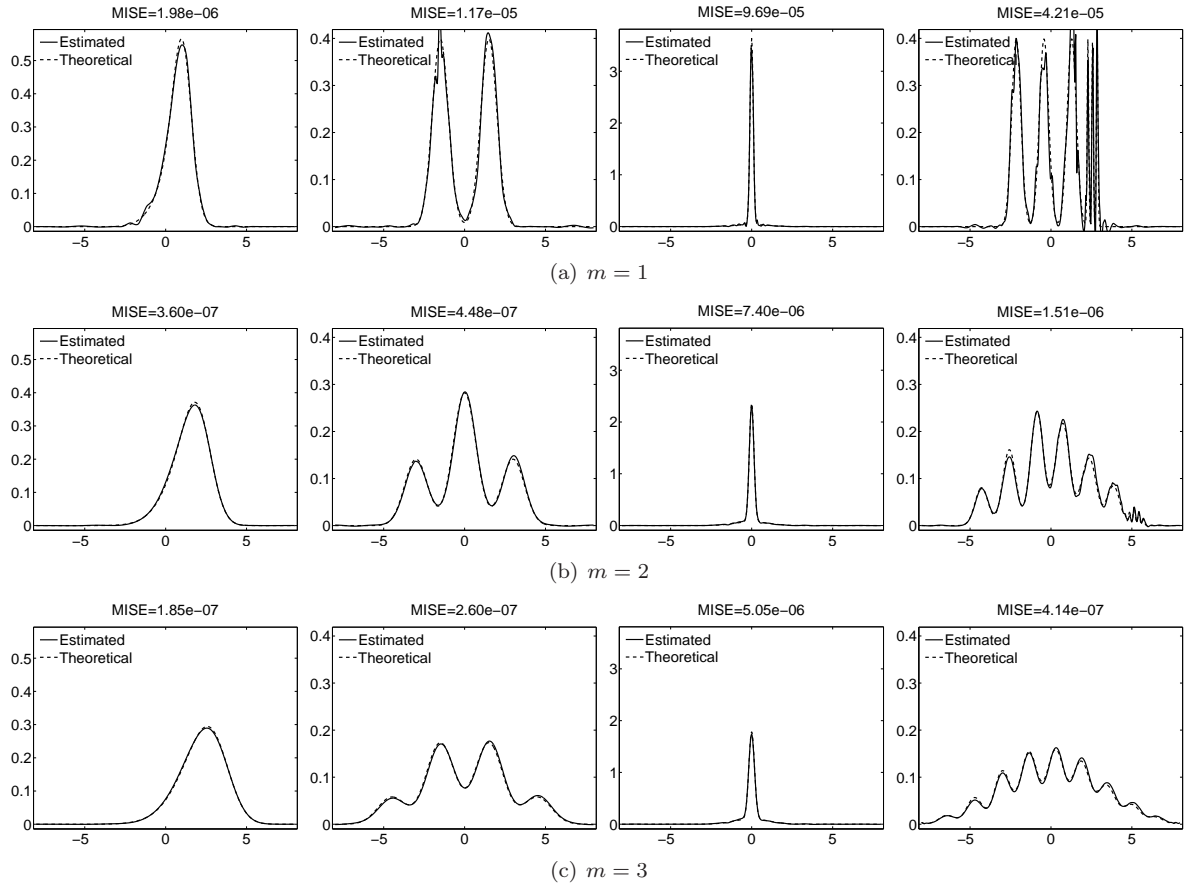


Figure 2: Original densities (dashed) and our wavelet hard thresholding estimator \hat{g} (solid) from only one repetition of $n = 1000$ samples Z_1, \dots, Z_n . From left to right SkewedUnimodal, SeparatedBimodal, Outlier and DiscreteComb. (a): $m = 1$, (b): $m = 2$, (c): $m = 3$.

where $\hat{\sigma}$ is the sample standard deviation and Q is the interquartile range. This choice was motivated by the major drawback of the Frees estimator which is the computation time required to evaluate it (see Table 2).

We evaluated the three procedures on small to medium sample. Each method was applied to $n = 10, 20, 50, 100$ data points of each of the densities. All experiments were conducted using a Gaussian kernel for both kernel-based methods. The MISE from 100 repetitions are tabulated in Table 1. It shows that none of the methods clearly outperforms the others in all cases. However, our estimator is slightly better than the others in many cases. Table 2 reports the average execution times in seconds for $m = 2$ and $m = 3$. For the Frees estimator, the computational cost increases dramatically as far as the sampling parameter n increases and during the computation to estimate the density of the sum of more than two *i.i.d.* random variables. From a practical point of view, unlike Frees's estimator, methods based on a plug-in approach can easily be computed for $m > 3$ and larger samples.

We conclude this section by a comparison to the natural kernel plug-in estimator of Saavedra and Cao (2000) on larger samples ($n = 1000, 2000, 5000$). Table 3 summarizes the results. Our wavelet method clearly outperforms the kernel one for all tests densities and all sample.

5 Application II: the deconvolution density model

5.1 Upper bound

We observe n *i.i.d.* random variables Z_1, \dots, Z_n where, for any $i \in \{1, \dots, n\}$,

$$Z_i = X_i + \epsilon_i, \quad (5.1)$$

X_1, \dots, X_n are *i.i.d.* random variables and $\epsilon_1, \dots, \epsilon_n$ are *i.i.d.* random variables. Classically, X_1, \dots, X_n are measurements of some characteristic of interest contaminated by noise represented by $\epsilon_1, \dots, \epsilon_n$. For

Table 1: $1e4 \times$ MISE values from 100 replications for each method. From top to bottom SkewedUnimodal, SeparatedBimodal, Outlier, AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right).

1.0e-04×	<i>SkewedUnimodal, m = 2</i>				<i>SkewedUnimodal, m = 3</i>				
n	10	20	50	100	n	10	20	50	100
Wavelet	89.52	32.10	11.47	6.90	Wavelet	73.02	28.63	10.44	6.44
Kernel	55.76	25.78	10.64	6.97	Kernel	52.16	23.88	9.79	6.49
Frees	80.56	32.09	11.52	7.08	Frees	85.79	31.78	10.66	6.60
<i>SeparatedBimodal, m = 2</i>					<i>SeparatedBimodal, m = 3</i>				
n	10	20	50	100	n	10	20	50	100
Wavelet	79.17	38.08	12.30	6.32	Wavelet	52.36	27.10	8.60	4.39
Kernel	57.97	45.59	36.57	35.20	Kernel	36.65	25.25	15.87	14.31
Frees	60.45	33.23	13.01	7.03	Frees	54.06	26.52	8.80	4.41
<i>Outlier, m = 2</i>					<i>Outlier, m = 3</i>				
n	10	20	50	100	n	10	20	50	100
Wavelet	542.4	257.4	82.61	49.81	Wavelet	525.6	251.9	81.27	48.37
Kernel	597.1	248.1	88.78	57.39	Kernel	750.8	264.4	84.16	53.29
Frees	602.8	265.6	86.42	49.87	Frees	691.8	271.2	84.81	49.41
<i>DiscreteComb, m = 2</i>					<i>DiscreteComb, m = 3</i>				
n	10	20	50	100	n	10	20	50	100
Wavelet	87.35	42.13	14.20	7.40	Wavelet	46.14	21.72	7.36	3.86
Kernel	44.62	31.72	26.51	24.55	Kernel	27.03	14.97	9.65	7.55
Frees	50.74	32.05	20.92	14.23	Frees	37.25	16.87	6.28	3.40

Table 2: Execution times in seconds for $m = 2$ and $m = 3$ (from only one realization). The algorithms were run under Matlab with an Intel Core 2 duo 3.06GHz CPU, 4Gb RAM.

$m = 2$					$m = 3$					
n	10	20	50	100	n	10	20	50	100	200
Wavelet	0.01	0.01	0.01	0.02	Wavelet	0.01	0.01	0.01	0.02	0.02
Kernel	0.04	0.04	0.04	0.05	Kernel	0.04	0.04	0.04	0.05	0.05
Frees	0.17	0.22	0.53	1.38	Frees	0.37	1.30	9.62	143	2037

any $i \in \{1, \dots, n\}$, X_i and ϵ_i are independent. The density of X_1 is unknown and denoted f , whereas the one of ϵ_1 is known and denoted h . For a fixed integer $m \geq 2$, let $S = \sum_{i=1}^m X_i$ and g be the density of S . The goal is to estimate g from Z_1, \dots, Z_n . This problem can be viewed as a generalization of the standard deconvolution density one which corresponds to $m = 1$. See e.g. [Caroll and Hall \(1988\)](#), [Fan \(1991\)](#), [Fan and Liu \(1997\)](#), [Pensky and Vidakovic \(1999\)](#), [Fan and Koo \(2002\)](#), [Butucea and Matias \(2005\)](#), [Comte et al. \(2006\)](#), [Delaigle and Gijbels \(2006\)](#) and [Lacour \(2006\)](#). However, to the best of our knowledge, the general problem i.e. with $m \geq 2$ is a new challenge.

Proposition 5.1 below investigates the rates of convergence of \hat{g} (2.7) constructed from a specific wavelet hard thresholding estimator \hat{f} under the MISE over Besov balls.

Proposition 5.1. Consider (5.1) and the associated notations. We define the Fourier transform of an integrable function u by $\mathcal{F}(u)(x) = \int_{-\infty}^{\infty} u(y)e^{-ixy}dy$, $x \in \mathbb{R}$. The notation $\overline{\cdot}$ will be used for the complex conjugate.

Suppose that $\text{supp } f \subseteq [-T, T]$, where $T > 0$ is a fixed constant, and there exist three constants $C > 0$, $c > 0$ and $\delta > 1$ such that

$$\sup_{x \in \mathbb{R}} h(x) \leq C, \quad |\mathcal{F}(h)(x)| \geq \frac{c}{(1+x^2)^{\delta/2}}, \quad x \in \mathbb{R}. \quad (5.2)$$

Let \hat{g} be (2.7) with

$$\hat{\alpha}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^n \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\phi_{j,k})(x)}}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx \quad (5.3)$$

Table 3: $1e6 \times$ MISE values from 100 replications. From top to bottom SkewedUnimodal, SeparatedBimodal, Outlier and AsymmetricClaw for $m = 2$ (left) and $m = 3$ (right).

$1.0e-06 \times$	<i>SkewedUnimodal, m = 2</i>			<i>SkewedUnimodal, m = 3</i>			
n	1000	2000	5000	n	1000	2000	5000
Wavelet	25.32	5.62	2.90	Wavelet	24.35	5.24	2.30
Kernel	27.46	10.09	4.86	Kernel	26.16	8.34	3.53
	<i>SeparatedBimodal, m = 2</i>			<i>SeparatedBimodal, m = 3</i>			
n	1000	2000	5000	n	1000	2000	5000
Wavelet	19.94	9.12	4.02	Wavelet	14.26	7.75	2.83
Kernel	509.5	394.9	258.8	Kernel	224.5	185.3	129.3
	<i>Outlier, m = 2</i>			<i>Outlier, m = 3</i>			
n	1000	2000	5000	n	1000	2000	5000
Wavelet	148.1	39.62	21.46	Wavelet	139.5	37.05	24.18
Kernel	182.6	54.94	34.86	Kernel	153.6	45.32	32.97
	<i>DiscreteComb, m = 2</i>			<i>DiscreteComb, m = 3</i>			
n	1000	2000	5000	n	1000	2000	5000
Wavelet	16.77	10.64	3.68	Wavelet	9.48	4.90	1.74
Kernel	541.8	521.0	474.1	Kernel	125.3	118.6	111.2

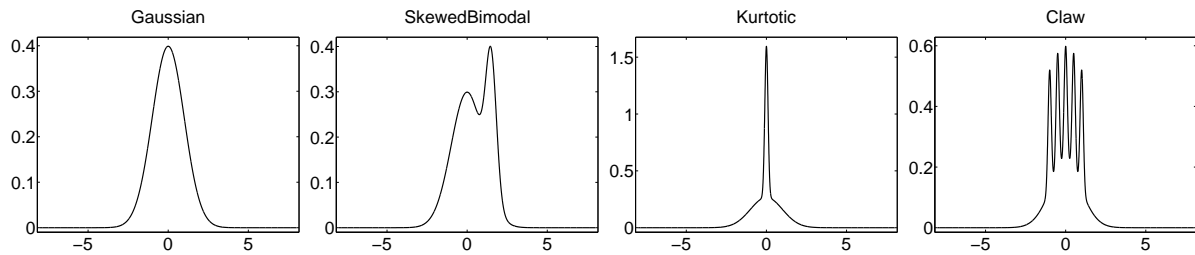


Figure 3: Theoretical densities.

and

$$\hat{\beta}_{j,k} = \frac{1}{2\pi n} \sum_{i=1}^n \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\psi_{j,k})(x)}}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx. \quad (5.4)$$

Suppose that $f \in B_{p,r}^s(M)$ with $r \geq 1$, $\{p \geq 4$ and $s > 0\}$ or $\{p \in [1, 4)$ and $s > \max((2\delta + 1)/p, (4/p - 1)(\delta + 1/2))\}$. Then there exists a constant $C > 0$ such that

$$\mathbb{E} \left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s/(2s+2\delta+1)}.$$

To the best of our knowledge, there is no asymptotic result for kernel estimators in this deconvolution setting. Proposition 5.1 provides a first theoretical result on the possible achievable rate of convergence for an estimator of g under the MISE over Besov balls. Let us mention that the rate of convergence $(\ln n/n)^{2s/(2s+2\delta+1)}$ corresponds to the ‘‘near optimal’’ one in the minimax sense for \hat{f} under the MISE over Besov balls. See (Fan and Koo, 2002, Theorem 2).

5.2 Simulation results

In this simulation, $n = 1000$ samples Z_1, \dots, Z_n were generated according to model (5.1) and we considered Laplace errors (which respect the standard ordinary smooth assumption). The data sets used in this deconvolution study are also normal mixture densities (see Marron and Wand (1992) for formulae of these densities) different representing degrees of smoothness.

Fig. 4 shows the results of \hat{g} for $m = 1$, $m = 2$ and $m = 3$ respectively. Clearly, for these nine densities, even if the estimation problem becomes harder our adaptive hard thresholding estimator is very effective.

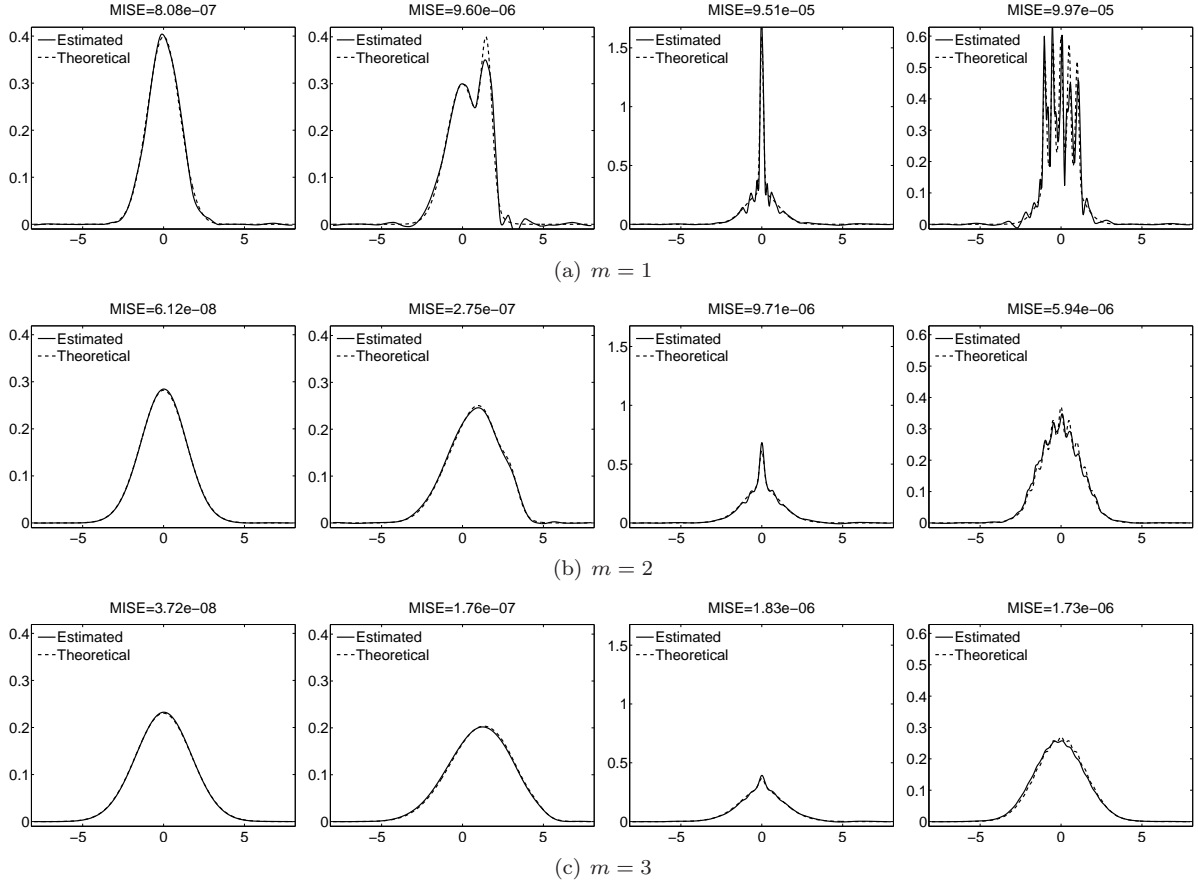


Figure 4: Original densities (dashed) and our wavelet hard thresholding estimator \hat{g} (solid) from one realization of $n = 1000$ observations Z_1, \dots, Z_n generated according to (5.1), where $h(x) = (1/2)e^{-|x|}$, $x \in \mathbb{R}$. (a): $m = 1$. (b): $m = 2$. (c): $m = 3$. (note that (5.2) is satisfied with $\delta = 2$).

Conclusion and perspectives

The agreement of our simulations with our theoretical findings show the relevance of our estimator in the context of two classical density estimation problems. The practical comparisons to state-of-the-art methods such as the estimator of [Frees \(1994\)](#) or the one of [Saavedra and Cao \(2000\)](#) have demonstrated the usefulness and the efficiency of adaptive thresholding methods in estimating densities of the sum of random variables. It would be interesting to include both theoretical and practical comparisons with other wavelet thresholding estimators as the block thresholding one (see e.g. [Cai \(1999\)](#) and [Chesneau et al. \(2010\)](#)). Another theoretical challenge is to determine the optimal lower bounds under the MISE over Besov balls. These aspects need further investigations that we leave for a future work.

6 Proofs

Proof of Theorem 3.1. Let us define the Fourier transform of a function $u \in \mathbb{L}_1(\mathbb{R})$ by

$$\mathcal{F}(u)(y) = \int_{-\infty}^{\infty} u(x)e^{-iyx} dx, \quad y \in \mathbb{R}.$$

By definition of \hat{g} and g , we have $\mathcal{F}(\hat{g})(y) = \mathcal{F}(\star^m \hat{f})(y) = (\mathcal{F}(\hat{f})(y))^m$ and $\mathcal{F}(g)(y) = \mathcal{F}(\star^m f)(y) = (\mathcal{F}(f)(y))^m$. Owing to the previous equalities and the Parseval theorem, we obtain

$$\int |\hat{g}(x) - g(x)|^2 dx = \frac{1}{2\pi} \int |(\mathcal{F}(\hat{f})(y))^m - (\mathcal{F}(f)(y))^m|^2 dy. \quad (6.1)$$

Now remark that, for any $(u, v) \in \mathbb{C}^2$, the factor theorem yields: $u^m - v^m = (u - v) \sum_{k=0}^{m-1} v^k u^{(m-1)-k}$. It follows from $\binom{m-1}{k} \geq 1$, $k \in \{0, \dots, m-1\}$, and the binomial theorem that

$$\begin{aligned} |u^m - v^m| &\leq |u - v| \sum_{k=0}^{m-1} |v|^k |u|^{(m-1)-k} \leq |u - v| \sum_{k=0}^{m-1} \binom{m-1}{k} |v|^k |u|^{(m-1)-k} \\ &= |u - v| (|u| + |v|)^{m-1} \leq |u - v| (|u - v| + 2|v|)^{m-1}. \end{aligned} \quad (6.2)$$

Using (6.1), (6.2), the linearity of \mathcal{F} , the inequalities: $|\mathcal{F}(\widehat{f} - f)(y)| \leq \int |\widehat{f}(x) - f(x)| dx$, $|\mathcal{F}(f)(y)| \leq \int |f(x)| dx \leq C$, and the Parseval theorem, we have

$$\begin{aligned} &\int |\widehat{g}(x) - g(x)|^2 dx \\ &\leq \frac{1}{2\pi} \int |\mathcal{F}(\widehat{f})(y) - \mathcal{F}(f)(y)|^2 \left(|\mathcal{F}(\widehat{f})(y) - \mathcal{F}(f)(y)| + 2|\mathcal{F}(f)(y)| \right)^{2m-2} dy \\ &= \frac{1}{2\pi} \int |\mathcal{F}(\widehat{f} - f)(y)|^2 \left(|\mathcal{F}(\widehat{f} - f)(y)| + 2|\mathcal{F}(f)(y)| \right)^{2m-2} dy \\ &\leq C \left(\int |\widehat{f}(x) - f(x)| dx + 1 \right)^{2m-2} \int |\widehat{f}(x) - f(x)|^2 dx. \end{aligned}$$

Noticing that the last term is the product of two random variables, the Cauchy-Schwarz inequality yields

$$\mathbb{E} \left(\int |\widehat{g}(x) - g(x)|^2 dx \right) \leq CIJ, \quad (6.3)$$

where

$$I = \sqrt{\mathbb{E} \left(\left(\int |\widehat{f}(x) - f(x)| dx + 1 \right)^{4m-4} \right)}, \quad J = \sqrt{\mathbb{E} \left(\left(\int |\widehat{f}(x) - f(x)|^2 dx \right)^2 \right)}.$$

Let us now bound I and J , in turn.

Upper bound for I . Using $|x + y|^a \leq 2^{a-1}(|x|^a + |y|^a)$, $(x, y) \in \mathbb{R}^2$, $a \geq 1$, the Hölder inequality and $\text{supp } f \subseteq [-T, T]$, we have

$$I \leq C \sqrt{\mathbb{E} \left(\int |\widehat{f}(x) - f(x)|^{4m-4} dx \right) + 1}.$$

Using the definition of \widehat{f} (2.6) and the wavelet expansion of f we have

$$\mathbb{E} \left(\int |\widehat{f}(x) - f(x)|^{4m-4} dx \right) \leq C(E + F + G),$$

where

$$\begin{aligned} F &= \int \mathbb{E} \left(\left(\sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} (\widehat{\beta}_{j,k} \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa 2^{\delta_j} \sqrt{\ln n/n}\}} - \beta_{j,k}) \psi_{j,k}(x) \right)^{4m-4} \right) dx \\ E &= \sum_{k \in \Lambda_\tau} \mathbb{E} (|\widehat{\alpha}_{\tau,k} - \alpha_{\tau,k}|^{4m-4}), \quad G = \int \left(\sum_{j=j_1+1}^{\infty} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x) \right)^{4m-4} dx \end{aligned}$$

Using (2.4) and $\int |f(x)|^{4m-4} dx \leq C$, we obtain

$$E + G \leq C.$$

Now observe that

$$|\widehat{\beta}_{j,k} \mathbf{1}_{\{|\widehat{\beta}_{j,k}| \geq \kappa 2^{\delta_j} \sqrt{\ln n/n}\}} - \beta_{j,k}| \leq |\widehat{\beta}_{j,k} - \beta_{j,k}| + |\beta_{j,k}|,$$

The unconditional property of the wavelet basis (see (Kerkycharian and Picard, 2000, Subsection 4.2)) gives

$$F \leq C(G + H)$$

where

$$G = \int \mathbb{E} \left(\left(\sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} (\hat{\beta}_{j,k} - \beta_{j,k}) \psi_{j,k}(x) \right)^{4m-4} \right) dx, \quad H = \int \left(\sum_{j=\tau}^{j_1} \sum_{k \in \Lambda_j} \beta_{j,k} \psi_{j,k}(x) \right)^{4m-4} dx.$$

Thanks to $\int |f(x)|^{4m-4} dx \leq C$, we have

$$H \leq C.$$

Arguing similarly to (Kerkycharian and Picard, 2000, Theorem 5.1) and using (2.4), we obtain

$$G \leq C \left(\frac{\ln n}{n} \right)^{2m-2} 2^{(1+2\delta)j_1(2m-2)} \leq C.$$

Therefore

$$I \leq C. \tag{6.4}$$

Upper bound for J. It follows from the Hölder inequality and $\text{supp } f \subseteq [-T, T]$ that

$$J \leq C \sqrt{\mathbb{E} \left(\int |\hat{f}(x) - f(x)|^4 dx \right)}.$$

We now need the following result.

Theorem 6.1 (Kerkycharian and Picard (2000)). *Let $\theta > 1$ and \hat{f} be (2.6) under (2.3), (2.4) with $v = 2\theta$ and (2.5) with $(\ln n/n)^\theta$ instead of $(\ln n/n)^4$. Then, for any $r \geq 1$, any $\{p \geq \theta$ and $s > 0\}$ or any $\{p \in [1, \theta]$ and $s > \max((2\delta + 1)/p, (\theta/p - 1)(\delta + 1/2))\}$, there exists a constant $C > 0$ such that*

$$\sup_{f \in B_{p,r}^s(M)} \mathbb{E} \left(\int |\hat{f}(x) - f(x)|^\theta dx \right) \leq C \left(\frac{\ln n}{n} \right)^{\theta s / (2s + 2\delta + 1)}.$$

Theorem 6.1 can be proved using arguments similar to (Kerkycharian and Picard, 2000, Theorem 5.1) and (Chesneau, 2008, Theorem 4.2).

Thanks to Theorem 6.1 with $\theta = 4$, we have

$$J \leq C \left(\frac{\ln n}{n} \right)^{2s / (2s + 2\delta + 1)}. \tag{6.5}$$

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

$$\mathbb{E} \left(\int |\hat{g}(x) - g(x)|^2 dx \right) \leq C \left(\frac{\ln n}{n} \right)^{2s / (2s + 2\delta + 1)}.$$

Theorem 3.1 is proved. □

This ends the proof of Theorem 3.1. □

Proof of Proposition 4.1. Thanks to (Donoho et al., 1996, Subsection 5.1.1, (16) and (17)), under the assumptions $\text{supp } f \in [-T, T]$, the estimators $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$ (4.1) satisfy (2.3), (2.4) and (2.5) with $\delta = 0$. The proofs are based on the decomposition:

$$\hat{\beta}_{j,k} - \beta_{j,k} = \frac{1}{n} \sum_{i=1}^n U_i, \quad U_i = \psi_{j,k}(Z_i) - \beta_{j,k},$$

where U_1, \dots, U_n are i.i.d. with $\mathbb{E}(U_1) = 0$, $|U_1| \leq C 2^{j_1/2} \leq (n/\ln n)^{1/2}$ and $\mathbb{E}(U_1^2) \leq C$. Then (ii) follows from the Rosenthal inequality and (iii) from the Bernstein inequality. The point (i) is similar to (ii) but with ϕ instead of ψ . The rest of the proof follows from Theorem 3.1. □

Proof of Proposition 5.1. Owing to (Fan and Koo, 2002, E. Proof of Theorem 7), under the assumptions $\text{supp } f \subseteq [-T, T]$ and (5.2), the estimators $\hat{\alpha}_{j,k}$ (5.3) and $\hat{\beta}_{j,k}$ (5.4) satisfy (2.3), (2.4) and (2.5) with the same δ . The proofs are based on the decomposition:

$$\hat{\beta}_{j,k} - \beta_{j,k} = \frac{1}{n} \sum_{i=1}^n V_i, \quad V_i = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\overline{\mathcal{F}(\psi_{j,k})(x)}}{\mathcal{F}(h)(x)} e^{-ixZ_i} dx - \beta_{j,k},$$

where V_1, \dots, V_n are i.i.d. with $\mathbb{E}(V_1) = 0$, $|V_1| \leq C2^{j_1(\delta+1/2)} \leq (n/\ln n)^{1/2}$ and $\mathbb{E}(V_1^2) \leq C2^{2\delta j}$. Then (ii) follows from the Rosenthal inequality and (iii) from the Bernstein inequality. The point (i) is similar to (ii) but with ϕ instead of ψ . We obtain the desired result via Theorem 3.1.

□

Acknowledgements

This work is supported by ANR grant NatImages, ANR-08-EMER-009. We thanks Jalal Fadili for his suggestions which lead to the improved version of the paper.

References

- Ahmad, I.A. and Fan, Y. (2001), "Optimal bandwidth for kernel density estimators of functions of observations," *Statist. Probab. Lett.*, 51, 245–251.
- Ahmad, I.A. and Mugdadi, A.R. (2003), "Analysis of kernel density estimation of functions of random variables," *J. Nonparametric Statistics*, 15, 579–605.
- Antoniadis, A. (1997), "Wavelets in statistics: a review" (with discussion), *Journal of the Italian Statistical Society*, Series B, 6, 97–144.
- Butucea, C. and Matias, C. (2005), "Minimax estimation of the noise level and of the signal density in a semiparametric convolution model," *Bernoulli*, 11, 309–340.
- Cai T. (1999), "Adaptive wavelet estimation: a block thresholding and oracle inequality approach," *Annals of Statistics*, 27, 898–924.
- Caroll, R.J. and Hall, P. (1988), "Optimal rates of convergence for deconvolving a density," *J. Amer. Statist. Assoc.*, 83, 1184–1186.
- Chesneau, C. (2008), "Wavelet estimation via block thresholding : a minimax study under \mathbb{L}^p risk," *Statistica Sinica*, 18, 1007–1024.
- Chesneau C., Fadili M.J. and Starck J.-L. (2010), "Stein Block Thresholding For Image Denoising," *Applied and Computational Harmonic Analysis*, 28, 67–88.
- Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993), "Wavelets on the interval and fast wavelet transforms," *Applied and Computational Harmonic Analysis*, 24, 54–81.
- Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006), "Penalized contrast estimator for adaptive density deconvolution," *The Canadian Journal of Statistics*, 34, 431–452.
- Delaigle, A. and Gijbels, I. (2006), "Estimation of boundary and discontinuity points in deconvolution problems," *Statistica Sinica*, 16, 773–788.
- Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1996), "Density estimation by wavelet thresholding," *Annals of Statistics*, 24, 508–539.
- Du, J. and Schick, A. (2007), "Root-n consistency and functional central limit theorems for estimators of derivatives of convolutions of densities," *Internat. J. Statist. Management Systems*, 2, 67–87.
- Fan, J. (1991), "On the optimal rates of convergence for nonparametric deconvolution problem," *Ann. Statist.*, 19, 1257–1272.

- Fan, J. and Liu, Y. (1997), "A note on asymptotic normality for deconvolution kernel density estimators," *Sankhya*, 59, 138–141.
- Fan, J. and Koo, J.Y. (2002), "Wavelet deconvolution," *IEEE transactions on information theory*, 48, 734–747.
- Frees, E. (1994), "Estimating densities of functions of observations," *J. Amer. Statist. Assoc.*, 89, 17–525.
- Giné, E. and Mason, D.M. (2007), "On local U-statistic processes and the estimation of densities of functions of several sample variables," *Ann. Statist.*, 35, 1105–1145.
- Härdle, W., Kerkycharian, G., Picard, D. and Tsybakov, A. (1998), "Wavelet, Approximation and Statistical Applications," Lectures Notes in Statistics, New York 129, Springer Verlag.
- Kerkycharian, G. and Picard, D. (2000), "Thresholding algorithms, maxisets and well concentrated bases (with discussion and a rejoinder by the authors)," *Test*, 9, 283–345.
- Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris Ser. I Math.*, 342 (11), 877–882.
- Mallat, S. (2009), "A wavelet tour of signal processing", Elsevier/Academic Press, Amsterdam, third edition. The sparse way, With contributions from Gabriel Peyré.
- Marron, J.S. and Wand, M.P. (1992), "Exact Mean Integrated Squared Error," *Ann. Statist.*, 20, 712–736.
- Meyer, Y. (1992), "Wavelets and Operators," Cambridge University Press, Cambridge.
- Mugdadi, A.R. and Ahmad, I. (2004), "A Bandwidth Selection for Kernel Density Estimation of Functions of Random Variables," *Computational Statistics and Data Analysis*, 47, 49–62.
- Panjer, H.H. and Willmot, G.E. (1992), "Insurance Risk Models", Society of Actuaries, Schaumburg.
- Pensky, M. and Vidakovic, B. (1999), "Adaptive wavelet estimator for nonparametric density deconvolution," *Ann. Statist.*, 27, 2033–2053.
- Prakasa Rao, B.L.S. (2004), "Moment inequalities for supremum of empirical processes of U-statistic structure and application to density estimation," *J.Iran. Statist. Soc.*, 3, 59–68.
- Saavedra, A. and Cao, R. (2000), "On the estimation of the marginal density of a moving average process," *Canad. J. Statist.*, 28, 799–815.
- Silverman, B. W. (1986), "Density estimation: for statistics and data analysis," Chapman and Hall, London.
- Schick, A. and Wefelmeyer, W. (2004), "Root n consistent density estimators for sums of independent random variables," *J. Nonparametr. Statist.*, 16, 925–935.
- Schick, A. and Wefelmeyer, W. (2007), "Root n consistent density estimators of convolutions in weighted \mathbb{L}_1 -norms," *J. Statist. Plann. Inference*, 137, 1765–1774.
- Tsybakov, A. (2004). "Introduction à l'estimation nonparamétrique," Springer Verlag, Berlin.
- Vidakovic, B. (1999), "Statistical Modeling by Wavelets," John Wiley & Sons, Inc., New York, 384 pp.