

Fast nonparametric estimation for convolutions of densities

CHRISTOPHE CHESNEAU¹, FABIENNE COMTE² AND FABIEN NAVARRO^{1,3}

¹*LMNO, Université de Caen Basse-Normandie, Département de Mathématiques, UFR de Sciences, 14032 Caen Cedex, France. E-mail: christophe.chesneau@unicaen.fr.*

²*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints Pères, 75006 Paris, France. E-mail: fabienne.comte@parisdescartes.fr.*

³*GREYC CNRS-ENSICAEN-Université de Caen, 14050 Caen Cedex, France. E-mail: fabien.navarro@ensicaen.fr.*

March 20, 2013

Abstract

The present paper is concerned with the problem of estimating the convolution of densities. We propose an adaptive estimator based on kernel methods, Fourier analysis and the Lepski method. We study its \mathbb{L}_2 -risk properties. Fast and new rates of convergence are determined for a wide class of unknown functions. Numerical illustrations, on both simulated and real data, are provided to assess the performances of our estimator.

AMS 2000 Subject Classifications: 62G07, 62G20.

Keywords. Convolution of densities. Kernel estimation. Minimax rate. Nonparametric estimation.

1 Introduction

1.1 Problem statement and motivations

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 modeled as the sum of all claims on the individual policies. Therefore, probability density functions of sums of random variables are of particular interest. Typically, such functions are not available in a closed form. Hence, in order to compute functionals of sums of random variables, estimation methods are often used.

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, f the unknown density function of a random variable $X : \Omega \rightarrow \mathbb{R}$, n a positive integer, X_1, \dots, X_n a n *i.i.d.* sample of X , m a positive integer and g the m -fold convolution defined by

$$g(x) = \star_m f(x) = \int \dots \int f(x - x_2 - \dots - x_m) f(x_2) \dots f(x_m) dx_2 \dots dx_m. \quad (1.1)$$

Let us note that g is the density of the random variable $S = \sum_{v=1}^m X_v$. We aim to estimate g from X_1, \dots, X_n . Sums of random variables occur in many situations in insurance and finance, for example, such a sum appears when considering the aggregate claims $X_1 + \dots + X_m$ of an insurance portfolio during a certain reference period (e.g. a month or a year). From actuarial point of view, one is often interested in the density function of S and m could be interpreted as the expected number of claims in a specified period. Moreover, the random variable S arises naturally in reliability theory as the lifetime of a system of m identical components. A detailed application in the field of health insurance can be found in Panjer and Willmot (1992). Other applications and examples, are indicated in Frees (1994). Zhang (2007) considers the estimation of sums of functions of observable and unobservable variables and provides examples in the field of data confidentiality problems and network. In the case of dependence between the random variables X_v , Cherubini *et al.* (2011) study the problem of capital allocation between risks when the sum of losses is bounded.

1.2 Previous work

The most famous nonparametric procedures are those developed by Frees (1994) and Saavedra and Cao (2000). Frees (1994) constructed a kernel based estimator using the random variables $S_\sigma = \sum_{v=1}^m X_{\sigma(v)}$, where $\sigma \in \{\text{combinations of } m \text{ elements } (\sigma(1), \dots, \sigma(m)) \text{ of } \{1, \dots, n\}\}$. Saavedra and Cao (2000) explored another point of view: the plug-in estimator $\hat{g} = \star_m \hat{f}$, where \hat{f} denotes a standard kernel estimator for f . These two estimators enjoy good mean integrated squared error (global \mathbb{L}_2 -risk) and asymptotic normality properties (see, e.g., 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, Theorems 3 and 4) show that $\hat{g} = \star_m \hat{f}$ attains the parametric rate of convergence $1/n$ under the global \mathbb{L}_2 -risk (or mean integrated square error MISE) if $m \geq 2$, f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$ and the functions f'' and $f^{(4)}$ are integrable. More recently, Nickl (2007) and mainly Nickl (2009) proved convergence in law results for kernel or wavelet estimators of the convolution of two densities under rather weak assumptions, to which it is worth comparing ours.

1.3 Contributions and relation to prior work

In this study we introduce a new kernel estimator \hat{g}_h , where h denotes the bandwidth. It is based on Fourier analysis following the spirit of, e.g., Fan (1991) and Carroll and Hall (1988) for the standard deconvolution density problem “ $f \star k$ ”, where k denotes a known density. In the first part, we establish sharp upper bounds for the pointwise and global \mathbb{L}_2 -risks of our estimator. Rates of convergence are determined under mild assumptions on f ; we only suppose that f belongs to Sobolev balls with smoothness parameter $\beta > 0$. We show the influence of β on the performances of \hat{g}_h and the best possible bandwidth choice. In particular, for each risks, we determine a constant $v_m > 0$ such that our estimator attains the parametric rate of convergence $1/n$ for $\beta > v_m$. When $\beta \in (0, v_m)$, this result does not hold; we exhibit a new rate of convergence depending on β . To the best of our knowledge, this phenomenon was never shown before. Let us mention that \hat{g}_h is not adaptive since the best h depends on the unknown β . However, for real data applications, such prior knowledge is not available.

This point motivates a second part devoted to the adaptive estimation of g . Focusing on the global \mathbb{L}_2 -risk, we develop a “Lepski method” (see, e.g., Lepski (1990) and Goldenshluger and Lepski (2010)) constructing an efficient estimator \hat{h} of the best bandwidth h for \hat{g}_h , whatever the smoothness of g . This yields an adaptive estimator $\hat{g}_{\hat{h}}$. We then study its rates of convergence under the global \mathbb{L}_2 -risk and assuming that f belongs to Sobolev balls with smoothness parameter $\beta > 0$. To be more specific, we prove that, if $m \geq 2$ and $\beta > 1/(2m(m-1))$, $\hat{g}_{\hat{h}}$ attains the parametric rate of convergence $1/n$. When $\beta \in (0, 1/(2m(m-1))]$, its rate of convergence is deteriorated by β . These asymptotic results are very sharp since they are close to those obtained by \hat{g}_h with the best nonadaptive h . The only difference is an extra logarithmic term when $\beta \in (0, 1/(2m(m-1))]$. We illustrate its performances via a simulation study and a real-data example is also provided to illustrate the application of the proposed estimator in a realistic situation.

1.4 Paper organization

The rest of the paper is organized as follows. Section 2 introduces our estimation procedure. The results are presented in Section 3. Simulated examples as well as a real-data application are provided in Section 4. The proofs are postponed to Section 5.

2 Estimation procedure

2.1 Notations

For any $a \in \mathbb{R}$, we set $[a]$ the integer part of a and, for any $(a, b) \in \mathbb{R}^2$, we denote $a \wedge b = \min(a, b)$.

For any $p \geq 1$, we set

$$\mathbb{L}_p(\mathbb{R}) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}; \|f\|_p = \left(\int_{-\infty}^{\infty} |f(x)|^p dx \right)^{1/p} \right\}.$$

Assuming that $f \in \mathbb{L}_1(\mathbb{R})$, we define the Fourier transform of f by

$$f^*(t) = \int_{-\infty}^{\infty} f(x)e^{-ixt} dx, \quad t \in \mathbb{R}. \quad (2.1)$$

Remark 2.1 *Let us recall that, if $f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$, the Fourier inverse formula yields that f is the inverse Fourier transform of f^* and can be written as*

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixy} f^*(y) dy, \quad x \in \mathbb{R}. \quad (2.2)$$

Let $L > 0$ and $\alpha > 0$. We define the Sobolev space $\mathcal{S}(\alpha, L)$ by

$$\mathcal{S}(\alpha, L) = \left\{ f \in \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R}), \quad \int_{-\infty}^{\infty} (1+x^2)^\alpha |f^*(x)|^2 dx \leq L \right\}.$$

2.2 Estimator

We consider the kernel estimator for f :

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{v=1}^n K\left(\frac{x - X_v}{h}\right),$$

where K denotes the sinus cardinal kernel, i.e.,

$$K(x) = \frac{\sin(\pi x)}{\pi x}, \quad x \in \mathbb{R}, \quad (2.3)$$

and h is a positive real number (called the "bandwidth").

Remark 2.2 *Let us mention that other choices of kernel functions K are possible (see, e.g., Tsybakov (2004)). In this study we focus our attention on the sinus cardinal kernel because it is compactly supported in the Fourier domain. Indeed, we have $K^*(t) = \mathbf{1}_{[-\pi, \pi]}(t)$, $t \in \mathbb{R}$.*

In view of (2.1), a natural plug-in estimator for f^* is given by:

$$\hat{f}_h^*(t) = \int_{-\infty}^{\infty} \hat{f}_h(x) e^{-ixt} dx = \tilde{f}^*(t) \mathbf{1}_{[-\pi/h, \pi/h]}(t),$$

where \tilde{f}^* denotes the empirical Fourier transform of f , i.e.,

$$\tilde{f}^*(t) = \frac{1}{n} \sum_{v=1}^n e^{-itX_v}. \quad (2.4)$$

Using the Fourier inverse formula (2.2) and the standard convolution equality: $(\star_m f)^*(t) = (f^*(t))^m$, observe that

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\star_m f)^*(t) e^{itx} dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} (f^*(t))^m e^{itx} dt. \quad (2.5)$$

Another plug-in in (2.5) yields the following estimator for g :

$$\hat{g}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} (\tilde{f}^*(t))^m e^{itx} dt. \quad (2.6)$$

2.3 A selection method for h

Let us now develop a bandwidth selection based on the so-called ‘‘Lepski method’’ and the global \mathbb{L}_2 -risk of \hat{g}_h . First of all, let us define

$$\hat{g}_{h,h'}(x) = \frac{1}{2\pi} \int_{-\pi/(h \vee h')}^{\pi/(h \vee h')} (\tilde{f}^*(t))^m e^{itx} dt, \quad x \in \mathbb{R},$$

$$V(h) = \kappa \frac{2\pi(\log(n))^m}{hn^m}$$

with $\kappa > 0$, and

$$A(h) = \sup_{h' \in \mathcal{H}_n} (\|\hat{g}_{h,h'} - \hat{g}_{h'}\|_2^2 - V(h'))_+.$$

Then we consider the following estimator for the optimal bandwidth h of \hat{g}_h under the global \mathbb{L}_2 -risk:

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}_n} (A(h) + V(h)), \quad (2.7)$$

where

$$\mathcal{H}_n = \left\{ h_k, \frac{1}{h_k} = k \in \{1, 2, \dots, n^m - 1, n^m\} \right\}.$$

Using \hat{h} (2.7) and \hat{g}_h (2.6), we consider the plug-in estimator $\hat{g}_{\hat{h}}$ for g .

Further details about the Lepski method can be found in, e.g., Lepski (1990), Golden-shluger and Lepski (2010) and Comte and Genon-Catalot (2012).

3 Results

3.1 Performances of \hat{g}_h under the pointwise \mathbb{L}_2 -risk

Let us start by the following remark.

Remark 3.1 *If f is a density such that $f \in \mathcal{S}(\beta, L)$, then $\sup_{u \in \mathbb{R}} (1 + u^2)^\beta |f^*(u)|^2 := B < +\infty$ (since $u \mapsto (1 + u^2)^\beta |f^*(u)|^2$ is continuous and integrable over \mathbb{R}), and*

$$\int (1 + u^2)^{m\beta} |g^*(u)|^2 du \leq B^{m-1} \int (1 + u^2)^\beta |f^*(u)|^2 du < B^{m-1} L,$$

so that $g \in \mathcal{S}(\alpha, M)$ with $\alpha = m\beta$ and $M = B^{m-1}L$. As m is fixed, we can still consider that α can be small (but then β is even smaller).

Proposition 3.1 below investigates the performance of \hat{g}_h under the pointwise \mathbb{L}_2 -risk.

Proposition 3.1 *Consider the model described in Section 1 and let \hat{g}_h be given by (2.6).*

Upper bound for the pointwise \mathbb{L}_2 -risk. Assume that $f^, g^* \in \mathbb{L}_1(\mathbb{R})$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{g}_h(x) - g(x))^2) \leq C \left(\frac{1}{h^2 n^m} + \frac{1}{n} + \left(\int_{|t| \geq \pi/h} |g^*(t)| dt \right)^2 \right).$$

Rate of convergence. Assume that $f^, g^* \in \mathbb{L}_1(\mathbb{R})$ and $f \in \mathcal{S}(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}((\hat{g}_h(x) - g(x))^2) \leq Cr_n, \tag{3.1}$$

where

Case	(i)	(ii)	(iii)
h	$m \geq 2, \beta \geq 1/m$	$m \geq 2, \beta \geq 1/(m(m-1))$	$\beta \in (0, 1/(m(m-1)))$
r_n	$O(n^{-1/2})$	$O(n^{-m/(2(m\beta+1))})$	$O(n^{-m/(2(m\beta+1))})$
	n^{-1}	n^{-1}	$n^{-m^2\beta/(m\beta+1)}$

In **(i)**, observe that \hat{g}_h is adaptive (since $h = O(n^{-1/2})$ does not depend on β) and attains the parametric rate of convergence $1/n$. The cases **(ii)** and **(iii)** are complementary. In **(ii)**, \hat{g}_h is non-adaptive but still attains $1/n$. In **(iii)**, $1/n$ does not hold and \hat{g}_h is still non-adaptive.

Note that we recover the regularity threshold $1/2$ when $m = 2$, that is given in Nickl (2009), see his Propositions 4 and 5. But Nickl (2009) studies convergence in law in $C_0(\mathbb{R})$, the space of bounded continuous functions on \mathbb{R} that vanish at infinity, while we only consider here pointwise risk. On the other hand, we find out that the regularity threshold for the parametric rate is $1/(m(m-1))$ and obtain nonparametric rates below the threshold, which is new.

3.2 Performances of \hat{g}_h

Proposition 3.2 below investigates the performance of \hat{g}_h under the global \mathbb{L}_2 -risk.

Proposition 3.2 *Consider the model described in Section 1 and let \hat{g}_h be defined by (2.6).*

Upper bound for the global \mathbb{L}_2 -risk. Assume that $g^ \in \mathbb{L}_2(\mathbb{R})$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right).$$

Rates of convergence. Assume that $f \in S(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq Cr_n,$$

where

Case	(iv)	(v)	(vi)
	$m \geq 2, \beta \geq 1/(2m)$	$m \geq 2, \beta \geq 1/(2m(m-1))$	$\beta \in (0, 1/(2m(m-1)))$
h	$O(n^{-1})$	$O(n^{-m/(2m\beta+1)})$	$O(n^{-m/(2m\beta+1)})$
r_n	n^{-1}	n^{-1}	$n^{-2m^2\beta/(2m\beta+1)}$

The proof of Proposition 3.2 is based on a suitable decomposition of the global \mathbb{L}_2 -risk, moments inequality for (2.4) and technical elements related to the Fourier analysis.

As in Proposition 3.1, Proposition 3.2 shows that, in **(iv)**, the procedure \hat{g}_h is adaptive and attains the parametric rate of convergence $1/n$. The cases **(v)** and **(vi)** are complementary. In **(v)**, \hat{g}_h is non-adaptive but still attains $1/n$. In **(vi)**, the rate of convergence $1/n$ is deteriorated and \hat{g}_h is still non-adaptive.

When $m = 1$ is considered in **(vi)**, the rate of convergence becomes the standard one for the density estimation problem via kernel method, i.e., $r_n = n^{-2\beta/(2\beta+1)}$ (see, e.g., Tsybakov (2004)).

Note that the results of Proposition 3.2 on the integrated error exhibit rates and bandwidth choices which are different from pointwise ones, under weak conditions. For instance, the parametric rate of convergence $1/n$ for the MISE is also attained by the kernel method of Saavedra and Cao (2000) (see (Saavedra and Cao, 2000, Theorems 3 and 4)). However, it is established with more restrictive conditions on f , compared to **(iv)**, i.e., f four times differentiable with fourth derivative continuous, $\sup_{x \in \mathbb{R}} |f^{(j)}(x)| < \infty$ for $j \in \{0, 1, 2, 3, 4\}$ and the functions f'' and $f^{(4)}$ are integrable. Nevertheless, similar results, under weak conditions, may be deduced from the main Theorem in Nickl (2009).

Naturally, the performances of \hat{g}_h deeply depends on h . We see in Propositions 3.1 and 3.2 that its optimal value depends on the unknown smoothness parameter β . In order to estimate it efficiently via the observations, a selection method was proposed in Section 2.3. The next section is devoted to the performances of the resulting adaptive estimator $\hat{g}_{\hat{h}}$.

3.3 Performances of $\hat{g}_{\hat{h}}$

Theorem 3.1 below explores the performances of $\hat{g}_{\hat{h}}$ under the global \mathbb{L}_2 -risk.

Theorem 3.1 *Consider the model described in Section 1. Let $\hat{g}_{\hat{h}}$ be defined by (2.6) with $h = \hat{h}$ from (2.7).*

Upper bound for the global \mathbb{L}_2 -risk. Assume that $g^ \in \mathbb{L}_2(\mathbb{R})$. Then there exist two constants κ and $C > 0$ such that*

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq C \left(\inf_{h \in \mathcal{H}_n} \left(\frac{1}{hn^m} + \int_{|t| \geq \pi/h} |g^*(t)|^2 + V(h) \right) + \frac{1}{n} \right).$$

Rates of convergence. Assume that $f \in S(\beta, L)$ for $\beta, L > 0$. Then there exists a constant $C > 0$ such that

$$\mathbb{E}(\|\hat{g}_{\hat{h}} - g\|_2^2) \leq Cr_n,$$

where

Case	(vii) $m \geq 2, \beta > 1/(2m(m-1))$	(viii) $\beta \in (0, 1/(2m(m-1))]$
r_n	n^{-1}	$\left(\frac{\log(n)}{n}\right)^{2m^2\beta/(2m\beta+1)}$

Theorem 3.1 shows that our adaptive procedure $\hat{g}_{\hat{h}}$ attains similar rates of convergence to those attained by \hat{g}_h with the optimal non-adaptive h . The only difference is a logarithmic loss for $\beta \in (0, 1/(2m(m-1))]$.

4 Numerical experiments

For simplicity in the following simulation study, we have added the superscript m on g to denote the m -th convolution power. We consider the problem of estimating the density g_m of the sum of a fixed number m of *i.i.d.* random variables, emphasizing the case $m = 2$. We demonstrate the usefulness of the adaptive kernel estimator $\hat{g}_{m, \hat{h}}$ on both simulated and real data examples. The numerical experiments have been carried out using Matlab.

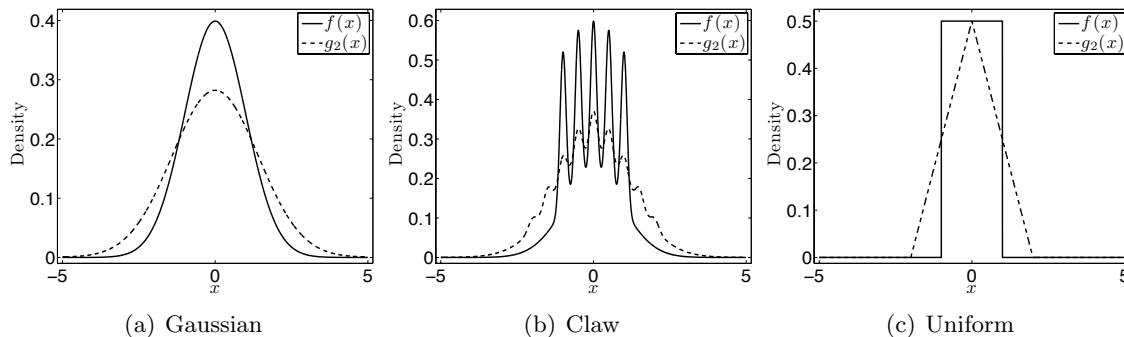


Figure 1: Test densities f (solid) and g_2 (dashed) (the 2-fold convolution power of f).

4.1 Computational aspects

A computationally-efficient procedure for kernel density estimation has been given by Silverman (1982, 1986), with extensions by Jones and Lotwick (1984). The main idea of this approach is to express kernel estimator as a convolution of the data with the kernel and use a Fast Fourier Transform (FFT) to perform the convolution (see e.g. Silverman (1986) Section 3.5). The estimator proposed in this paper is based on the same idea, therefore FFT-based methods are well suited for its implementation. The resulting estimator $\hat{g}_{m,h}$ is simple to implement and fast which allows us to perform the selection procedure in a reasonable time for various sample sizes.

The original observations were generated from an *i.i.d.* sample of random variables X_1, \dots, X_n . Three test functions, representing different degrees of smoothness were used (see Figure 1),

- (a) the standard normal distribution $\mathcal{N}(0, 1)$,
- (b) the Claw distribution $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=0}^4 \mathcal{N}(l/2 - 1, (1/2)^2)$,
- (c) the uniform distribution $\mathcal{U}(-1, 1)$.

For numerical implementation, we consider an interval $[a, b]$ that covers the range of the data and the density estimates were evaluated at $M = 2^r$ equally spaced points $t_i = a + (b - a)/M, i = 0, 1, \dots, M - 1$, between a and b , with $r = 8, b = -a = 5$ and M is the number of discretization points. The constant κ is taken equal to 1 and the normalized sinc Kernel (2.3) were used throughout all experiments. Our results are based on $N = 500$ simulated data sets and the MISE was approximated by the average of the Integrated Squared Error (ISE) over the N replicates.

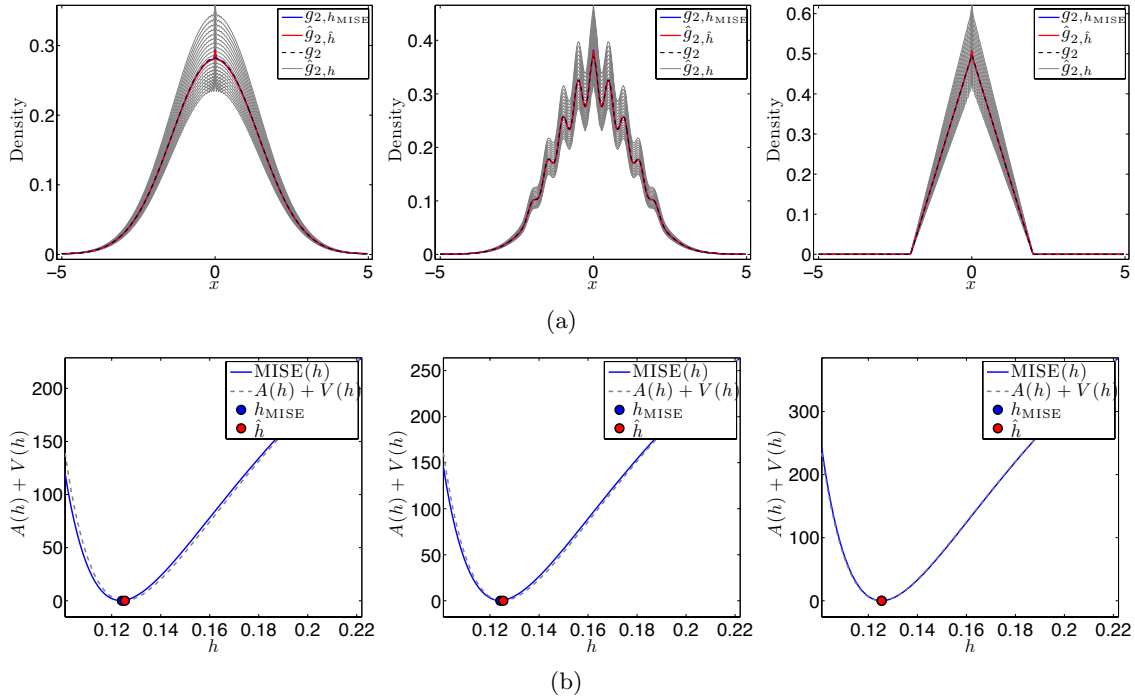


Figure 2: (a): True density (dotted), density estimates (red) and sample of 20 estimates (gray) out of 100 proposed to the selection algorithm obtained with a sample of $n = 1000$ data. (b): Graph of the the function $A(h) + V(h)$ against the smoothing parameter h and (shifted) $\text{MISE}(h)$. The red circle represents the global minimizer of $A(h) + V(h)$ and the blue circle represents the global minimizer of $\text{MISE}(h)$.

4.2 Bandwidth selection procedure

We have considered the case where $m = 2$ to illustrate the finite sample behavior of the adaptive estimator $\hat{g}_{m,\hat{h}}$, constructed with the data-driven bandwidth selection procedure described in the Section 2. Note that, although we focused on $m = 2$ the approach can easily handle other values of m .

In order to reduce, the computational cost of the selection procedure and improve the numerical efficiency of the resulting estimator, we have reduced the cardinality of the finite set of bandwidth \mathcal{H}_n . Indeed, a proper choice of an initial grid \mathcal{H}_n allows to reduce considerably the computational time involved in the bandwidth selection procedure and consequently in the estimation of $\hat{g}_{m,\hat{h}}$. Thus, in each case, the grid of h values that we have considered consisted of 100 values from $0.2\hat{h}_0$ to $1.2\hat{h}_0$ where $\hat{h}_0 = n^{-m/5}$ denotes a pilot bandwidth generally selected from a reference rule, like Silverman's rule of thumb (see e.g.

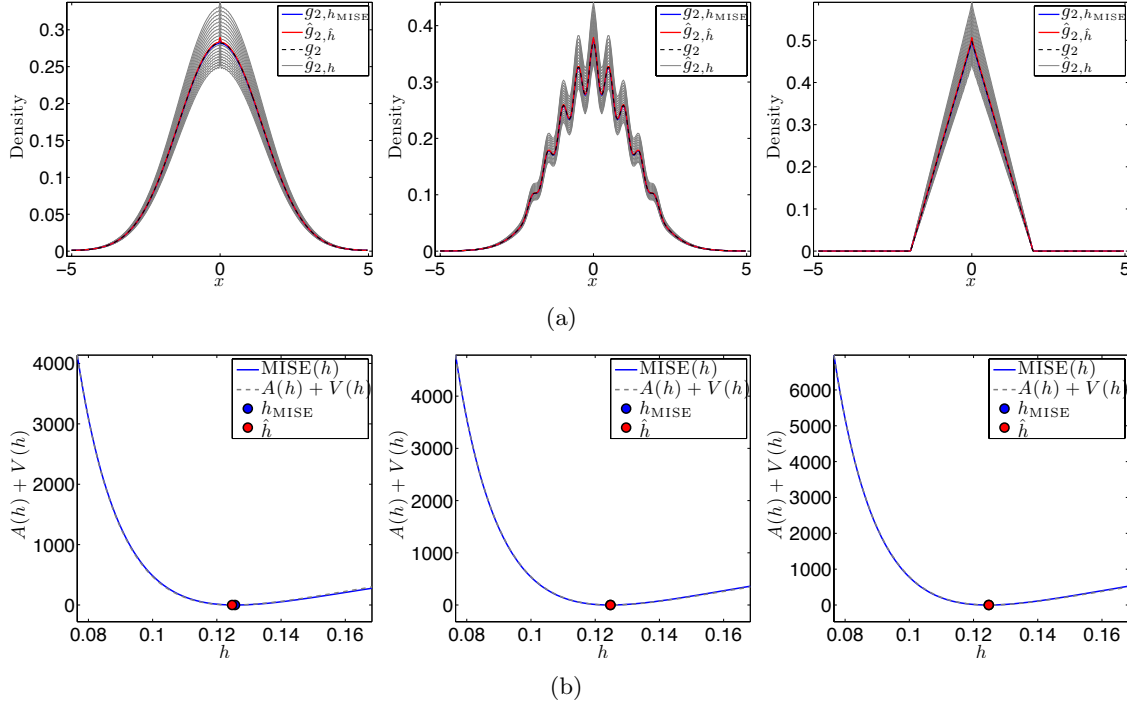


Figure 3: (a): True density (dotted), density estimates (red) and sample of 20 estimates (gray) out of 100 proposed to the selection algorithm obtained with a sample of $n = 2000$ data. (b): Graph of the function $A(h) + V(h)$ against the smoothing parameter h and (shifted) $\text{MISE}(h)$. The red circle represents the global minimizer of $A(h) + V(h)$ and the blue circle represents the true MISE -minimizing bandwidth h_{MISE} .

Silverman (1986)). Thus, in practice, for each value of $h \in \mathcal{H}_n$, the function $A(h) + V(h)$ has been computed from a single simulation and then, we have minimized (2.7) numerically over the grid \mathcal{H}_n . This step can be easily and quickly computed (thanks to the FFT-all algorithm). The results are depicted in Figures 2 and 3 for n -samples of sizes 1000 et 2000. Figures 2(b) 3(b) also contain a plot of the function $A(h) + V(h)$ against the smoothing parameter h and a vertical shift of the curve $\text{MISE}(h)$ is also overlayed for visualization purposes. For each density, it is clear from the figures, that the value of \hat{h} is the unambiguous minimizer of $A(h) + V(h)$. One notes that selection is already very effective for $n = 1000$, \hat{h} provides a decent approximation, fairly close to h_{MISE} and even closer when the sample size n grows (i.e. $n = 2000$), for all test densities. When $n = 1000$, for the Gaussian density, the bandwidth which minimizes $\text{MISE}(h)$ in this case is $h_{\text{MISE}} = 0.12520$ and $\hat{h} = 0.12543$. In this case, for the Claw and the uniform densities, we obtained respectively $h_{\text{MISE}} = 0.12517, \hat{h} = 0.12540!$ and $h_{\text{MISE}} = 0.12547, \hat{h} = 0.12542$. Note that,

the selection procedure also give good results on smaller sample sizes (e.g. for $n = 100$ and the Gaussian density, $h_{\text{MISE}} = 0.3980$ and $\hat{h} = 0.3888$). Therefore, without any prior smoothness knowledge on the unknown density, our adaptive estimator is very effective to estimate each of the three densities.

Table 1: $10000 \times \text{MISE}$ values from 500 replications. From top to bottom Gaussian, Claw and Uniform densities for $m = 2$. The numbers in parenthesis show the performance of Saavedra and Cao (2000)

n	500	1000	2000
Gaussian	0.0186 (0.0345)	0.0065 (0.0188)	0.0028 (0.0101)
Claw	0.0291 (0.1535)	0.0114 (0.1261)	0.0056 (0.1114)
Uniform	0.0250 (0.1301)	0.0104 (0.0867)	0.0072 (0.0589)

Finally, we have compared the performance of our adaptive estimator to those of a different kernel-based estimator presented in Saavedra and Cao (2000), i.e. $\hat{g} = \star_m \hat{f}$, where $\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right)$, is a standard kernel estimator and K is a kernel function. All experiments were conducted using a Gaussian kernel and we have been focused on a global bandwidth selector: the *rule of thumb* (ROT) bandwidth selector (see e.g. Silverman (1986)). Thus, the optimal bandwidth is given by $h_{\text{ROT}} = 1.06 \min(\hat{\sigma}, Q/1.34)n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation and Q is the interquartile range. TABLE 1 summarizes the results. It shows in particular that our adaptive estimator systematically outperforms the estimator Saavedra and Cao (2000). As expected and predicted by our theoretical findings, the performance gets better as the sample sizes increases.

Real data example

In this section a real data application of the bandwidth selection method is given. In insurance, a typical example is the sum of insurance claims, where $(X_v)_{v=1, \dots, n}$ are individual insurance claims and $S = \sum_{v=1}^m X_v$ 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). In this illustration, we consider the hospital data example, reported and analyzed in Frees (1994) and we discuss the case $m = 2$. The data in Table 2 consist of measurements of the 1989 total charges for 33 patients at a Wisconsin Hospital

The bandwidth was also chosen by grid search minimization, over an equally spaced grid of 100 values from $h_{\min} = 9$ to $h_{\max} = 450$. Figure 4(a) depicts density estimate and show histogram of the sum of claims for $m = 2$. Figure 4(b) contain a plot of the function $A(h) + V(h)$ against the smoothing parameter h and the red circle represents the global

Table 2: 1989 total hospital charges (in dollars) for 33 females aged 30-49 hospitalized for circulatory disorders from a Wisconsin Hospital (see Frees (1994)).

2337	2179	2348	4765	2088	2872	1924	2294	2182	2138	1765
2467	3609	2141	1850	3191	3020	2473	1898	7787	6169	1802
2011	2270	3425	3558	2315	1642	5878	2101	2242	5746	3041

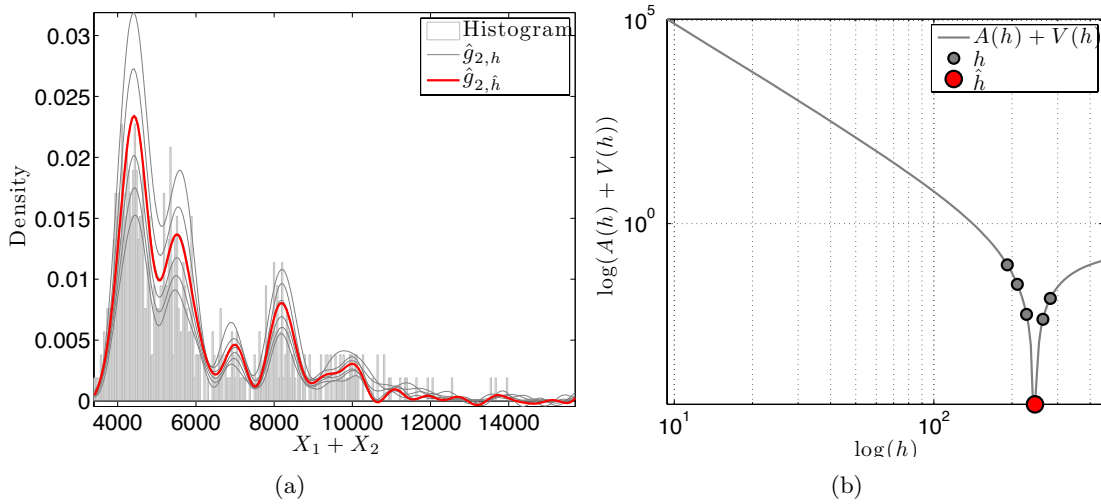


Figure 4: (a) Our estimator $\hat{g}_{2,\hat{h}}$ (red) for the sum of two claims and a sample of 4 estimates $\hat{g}_{2,h}$ (gray) out of 100 used to the selection algorithm. (b): Graph of $A(h) + V(h)$ against the smoothing parameter h (in a log-log scale). The red circle represents the global minimizer \hat{h} of $A(h) + V(h)$ and the gray circles represent the values of h corresponding to the \hat{g}_h plot in (a).

minimizer \hat{h} of $A(h) + V(h)$. The density estimates depicted by the gray lines in Figure 4(a) are based on the bandwidths marked by the gray circles in Figure 4(b).

Conclusion and perspectives

We have constructed a new adaptive estimator $\hat{g}_{\hat{h}}$ for g using kernel methods, Fourier analysis and the Lepski method. Theorems 3.2 and 3.1 show the good performances of $\hat{g}_{\hat{h}}$ in terms of rates of convergence under the global \mathbb{L}_2 -risk over $S(\beta, L)$. The agreement of our simulations with our theoretical findings show that our estimator is quite effective on

both simulated and real data sets. Possible perspectives of this work are to

- remove the extra logarithmic term in the case $\beta \in (0, 1/(2m(m-1))]$ via another adaptive estimator,
- extend our estimation procedure to another problem as the one of Zijaeva (1975), i.e., estimate the density g of $X_1 + Y_1$ from n *i.i.d.* observations X_1, \dots, X_n with density f_1 and n *i.i.d.* observations Y_1, \dots, Y_n with density f_2 ,

All these points need further investigations that we leave for future works.

5 Proofs

In the following, the quantity C denotes a generic constant that does not depend on n . Its value may change from one term to another (it can depend on the fixed m).

5.1 Intermediary results

Proposition 5.1 *Let f^* be (2.1) and \tilde{f}^* be (2.4). For any positive integer v , there exists a constant $C > 0$ such that*

$$\sup_{t \in \mathbb{R}} \mathbb{E} \left(|\tilde{f}^*(t) - f^*(t)|^{2v} \right) \leq C \frac{1}{n^v}.$$

Proof of Proposition 5.1. First of all, let us recall the Rosenthal inequality (see Rosenthal (1970)).

Lemma 5.1 (Rosenthal's inequality) *Let n be a positive integer, $p \geq 2$ and U_1, \dots, U_n be n zero mean independent random variables such that $\sup_{v \in \{1, \dots, n\}} \mathbb{E}(|U_v|^p) < \infty$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E} \left(\left| \sum_{v=1}^n U_v \right|^p \right) \leq C \left(\sum_{v=1}^n \mathbb{E}(|U_v|^p) + \left(\sum_{v=1}^n \mathbb{E}(U_v^2) \right)^{p/2} \right).$$

Let $t \in \mathbb{R}$. We have

$$\tilde{f}^*(t) - f^*(t) = \frac{1}{n} \sum_{v=1}^n U_v(t), \quad U_v(t) = e^{-itX_v} - \mathbb{E}(e^{-itX_1}).$$

Note that $U_1(t), \dots, U_n(t)$ are *i.i.d.* with $\mathbb{E}(U_1(t)) = 0$, $|U_1(t)| \leq 1$ and $\mathbb{E}(|U_1(t)|^2) \leq 1$. For any positive integer v , the Rosenthal inequality yields the existence of a constant $C > 0$ such that

$$\mathbb{E} \left(|\tilde{f}^*(t) - f^*(t)|^{2v} \right) = \frac{1}{n^{2v}} \mathbb{E} \left(\left| \sum_{v=1}^n U_v(t) \right|^{2v} \right) \leq C \frac{1}{n^v}.$$

Proposition 5.1 is proved. □

Lemma 5.2 *Let $(u, v) \in \mathbb{C}^2$ such that $|u| \leq 1$ and $|v| \leq 1$. Then, for any integer $m \geq 1$, we have*

$$|u^m - v^m| \leq |u - v|^m + D_m |u - v| |v|,$$

with $D_m = (3^m - 2^m - 1)/2$.

Proof of Lemma 5.2. For $m = 1$, the desired inequality is obviously satisfied with $D_m = 0$. Let us now investigate the case $m \geq 2$. The binomial theorem yields

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

The triangular inequality and $|u| \leq 1$ and $|v| \leq 1$ give

$$|u^m - v^m| \leq |u - v|^m + D_m |u - v| |v|,$$

with

$$D_m = 2^{m-2} \sum_{k=0}^{m-2} \binom{m}{k+1} 2^{-k} = \frac{1}{2}(3^m - 2^m - 1).$$

Lemma 5.2 is proved. □

5.2 Proof of the main results

5.2.1 Proof of Proposition 3.1

Upper bound for the pointwise L_2 risk. Set

$$\bar{g}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} g^*(t) e^{itx} dt = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} (f^*(t))^m e^{itx} dt, \quad x \in \mathbb{R}. \quad (5.1)$$

Using the elementary inequality $(x + y)^2 \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, we obtain

$$\mathbb{E}((\hat{g}_h(x) - g(x))^2) \leq 2(P + Q), \quad (5.2)$$

where

$$P = \mathbb{E} \left((\hat{g}_h(x) - \bar{g}_h(x))^2 \right), \quad Q = (\bar{g}_h(x) - g(x))^2.$$

Upper bound for Q. The Fourier inverse formula yields

$$Q = \left(\frac{1}{2\pi} \int_{|t| \geq \pi/h} g^*(t) e^{itx} dt \right)^2 \leq \left(\frac{1}{2\pi} \int_{|t| \geq \pi/h} |g^*(t)| dt \right)^2. \quad (5.3)$$

Upper bound for Q. We have

$$\begin{aligned} P &= \mathbb{E} \left(\left(\frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} ((\tilde{f}^*(t))^m - (f^*(t))^m) e^{itx} dt \right)^2 \right) \\ &\leq \mathbb{E} \left(\left(\frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} |(\tilde{f}^*(t))^m - (f^*(t))^m| dt \right)^2 \right) \\ &\leq \left(\frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \left(\mathbb{E} \left(|(\tilde{f}^*(t))^m - (f^*(t))^m|^2 \right) \right)^{1/2} dt \right)^2, \end{aligned} \quad (5.4)$$

where the last line follows from the Fubini theorem (write the squared integral as a multiple integral) and the Cauchy-Schwarz inequality.

It follows from the inequalities $|\tilde{f}^*(t)| \leq 1$, $|f^*(t)| \leq \|f\|_1 = 1$, Lemma 5.2 and the elementary inequality $(x + y)^2 \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, that

$$|(\tilde{f}^*(t))^m - (f^*(t))^m|^2 \leq 2 \left(|\tilde{f}^*(t) - f^*(t)|^{2m} + D_m^2 |\tilde{f}^*(t) - f^*(t)|^2 |f^*(t)|^2 \right),$$

with $D_m = (3^m - 2^m - 1)/2$.

Proposition 5.1 implies the existence of a constant $C > 0$ such that

$$\begin{aligned} &\mathbb{E} \left(|(\tilde{f}^*(t))^m - (f^*(t))^m|^2 \right) \\ &\leq 2 \left(\mathbb{E} \left(|\tilde{f}^*(t) - f^*(t)|^{2m} \right) + D_m^2 |f^*(t)|^2 \mathbb{E} \left(|\tilde{f}^*(t) - f^*(t)|^2 \right) \right) \\ &\leq C \left(\frac{1}{n^m} + \frac{1}{n} |f^*(t)|^2 \right). \end{aligned} \quad (5.5)$$

By (5.4), (5.5), the elementary inequalities $\sqrt{|x + y|} \leq \sqrt{|x|} + \sqrt{|y|}$, and $(x + y)^2 \leq$

$2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, and $\int_{-\pi/h}^{\pi/h} |f^*(t)| dt \leq \|f^*\|_1$, we have

$$\begin{aligned} P &\leq C \left(\int_{-\pi/h}^{\pi/h} \left(\frac{1}{n^m} + \frac{1}{n} |f^*(t)|^2 \right)^{1/2} dt \right)^2 \\ &\leq C \left(\frac{1}{n^m h^2} + \frac{1}{n} \left(\int_{-\pi/h}^{\pi/h} |f^*(t)| dt \right)^2 \right) \leq C \left(\frac{1}{n^m h^2} + \frac{1}{n} \right). \end{aligned} \quad (5.6)$$

Combining (5.2), (5.3) and (5.6), we obtain

$$\mathbb{E} ((\hat{g}_h(x) - g(x))^2) \leq C \left(\frac{1}{h^2 n^m} + \frac{1}{n} + \left(\int_{|t| \geq \pi/h} |g^*(t)| dt \right)^2 \right). \quad (5.7)$$

Rates of convergence. Since $\int_{-\infty}^{\infty} |f^*(t)|^2 (1+t^2)^\beta dt \leq L$, recalling that this implies that $|f^*(t)|^2 (1+t^2)^\beta \leq B$, we have

$$\begin{aligned} \left(\int_{|t| \geq \pi/h} |g^*(t)| dt \right)^2 &= \left(\int_{|t| \geq \pi/h} |f^*(t)|^m dt \right)^2 \\ &= \left(\int_{|t| \geq \pi/h} \left(|f^*(t)| (1+t^2)^{\beta/2} \right)^2 \left(|f^*(t)| (1+t^2)^{\beta/2} \right)^{m-2} (1+t^2)^{-m\beta/2} dt \right)^2 \\ &\leq (1 + (\pi/h)^2)^{-m\beta} B^{m-2} L^2 \leq C h^{2m\beta}. \end{aligned} \quad (5.8)$$

(i) $m \geq 2$, $\beta \geq 1/m$: The inequalities (5.7) and (5.8) and $h = O(n^{-1/2})$ give

$$\mathbb{E} ((\hat{g}_h(x) - g(x))^2) \leq C \left(\frac{1}{n^{m-1}} + \frac{1}{n} + n^{-m\beta} \right) \leq C \frac{1}{n}.$$

(ii) $m \geq 2$, $\beta \geq 1/(m(m-1))$: It follows from (5.7) and (5.8) and $h = O(n^{-m/(2(m\beta+1))})$ that

$$\mathbb{E} ((\hat{g}_h(x) - g(x))^2) \leq C \left(n^{-m^2\beta/(m\beta+1)} + \frac{1}{n} + n^{-m^2\beta/(m\beta+1)} \right) \leq C \frac{1}{n}.$$

(iii) $\beta \in (0, 1/(m(m-1)))$: Similarly to the previous case, we have

$$\mathbb{E} ((\hat{g}_h(x) - g(x))^2) \leq C \left(n^{-m^2\beta/(m\beta+1)} + \frac{1}{n} + n^{-m^2\beta/(m\beta+1)} \right) \leq C n^{-m^2\beta/(m\beta+1)}.$$

The proof of Proposition 3.1 is complete. \square

5.2.2 Proof of Proposition 3.2.

Upper bound for the global \mathbb{L}_2 -risk. Recall that \bar{g}_h is defined in equation (5.1). The elementary inequality $(x + y)^2 \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, yields

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq 2(R + S), \quad (5.9)$$

where

$$R = \mathbb{E} (\|\hat{g}_h - \bar{g}_h\|_2^2), \quad S = \|\bar{g}_h - g\|_2^2.$$

Upper bound for R. The Parseval theorem gives

$$R = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} \mathbb{E} (|(\tilde{f}^*(t))^m - (f^*(t))^m|^2) dt \quad (5.10)$$

Owing to the Parseval theorem, we have

$$\int_{-\pi/h}^{\pi/h} |f^*(t)|^2 dt \leq \|f^*\|_2^2 = 2\pi \|f\|_2^2 \leq C. \quad (5.11)$$

It follows from (5.10), (5.5) and (5.11) that

$$R \leq C \left(\frac{1}{hn^m} + \frac{1}{n} \int_{-\pi/h}^{\pi/h} |f^*(t)|^2 dt \right) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} \right). \quad (5.12)$$

Upper bound for S. Using the Parseval theorem, we get

$$S = \int_{|t| \geq \pi/h} |g^*(t)|^2 dt. \quad (5.13)$$

Combining (5.9), (5.12) and (5.13), we obtain

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right). \quad (5.14)$$

Rates of convergence. Let us recall that, since $f \in S(\beta, L)$, then $g \in S(\alpha, M)$ with $\alpha = m\beta$ and $M = B^{m-1}L$ (see Remark 3.1). Therefore

$$\int_{|t| \geq \pi/h} |g^*(t)|^2 dt \leq (1 + (\pi/h)^2)^{-m\beta} \int_{|t| \geq \pi/h} (1 + t^2)^{m\beta} |g^*(t)|^2 dt \leq CLh^{2m\beta}. \quad (5.15)$$

(iv) $m \geq 2$ and $\beta \geq 1/(2m)$: It follows from (5.14) and (5.15) with $h = O(n^{-1})$ that

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(\frac{1}{n^{m-1}} + \frac{1}{n} + n^{-2m\beta} \right) \leq C \frac{1}{n}.$$

(v) $m \geq 2$ and $\beta \geq 1/(2m(m-1))$: The choice $h = O(n^{-m/(2m\beta+1)})$ in the bounds (5.14) and (5.15) yields

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n} + n^{-2m^2\beta/(2m\beta+1)} \right) \leq C \frac{1}{n}.$$

(vi) $\beta \in (0, 1/(2m(m-1)))$: Proceeding as for the previous point, we obtain

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n} + n^{-2m^2\beta/(2m\beta+1)} \right) \leq C n^{-2m^2\beta/(2m\beta+1)}.$$

This ends the proof of Proposition 3.2. □

5.2.3 Proof of Theorem 3.1.

Upper bound for the global \mathbb{L}_2 -risk. Let $h \in \mathcal{H}_n$ be fixed. The Minkowski inequality and the elementary inequality $(x + y + z)^2 \leq 3(x^2 + y^2 + z^2)$, $(x, y, z) \in \mathbb{R}^3$, yield

$$\mathbb{E} (\|\hat{g}_{\hat{h}} - g\|_2^2) \leq 3(\mathbb{E} (\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\|_2^2) + \mathbb{E} (\|\hat{g}_{h,\hat{h}} - \hat{g}_h\|_2^2) + \mathbb{E} (\|\hat{g}_h - g\|_2^2)).$$

By definition of $A(h)$, we have

$$\mathbb{E} (\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\|_2^2) \leq \mathbb{E}(A(h)) + \mathbb{E}(V(\hat{h})),$$

the definition of $A(\hat{h})$ yields

$$\mathbb{E} (\|\hat{g}_{h,\hat{h}} - \hat{g}_h\|_2^2) \leq \mathbb{E}(A(\hat{h})) + V(h)$$

and Proposition 3.2 gives

$$\mathbb{E} (\|\hat{g}_h - g\|_2^2) \leq C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right).$$

Therefore, owing to the definition of \hat{h} , we have

$$\begin{aligned}
& \mathbb{E} (\|\hat{g}_{\hat{h}} - g\|_2^2) \\
& \leq 3(\mathbb{E}(A(h)) + \mathbb{E}(V(\hat{h}) + A(\hat{h})) + V(h)) + C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right) \\
& \leq 6(\mathbb{E}(A(h)) + V(h)) + C \left(\frac{1}{hn^m} + \frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right), \tag{5.16}
\end{aligned}$$

where

$$\mathbb{E}(A(h)) = \mathbb{E} \left(\sup_{h' \in \mathcal{H}_n} (\|\hat{g}_{h,h'} - \hat{g}_{h'}\|_2^2 - V(h'))_+ \right). \tag{5.17}$$

Upper bound for $\mathbb{E}(A(h))$. Let us introduce the following functions:

$$\bar{g}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} g^*(t) e^{itx} dt, \quad x \in \mathbb{R}$$

and

$$\bar{g}_{h,h'}(x) = \frac{1}{2\pi} \int_{-\pi(1/h \wedge 1/h')}^{\pi(1/h \wedge 1/h')} g^*(t) e^{itx} dt, \quad x \in \mathbb{R}.$$

Observe that

$$\hat{g}_{h'} - \hat{g}_{h,h'} = \hat{g}_{h'} - \bar{g}_{h'} - (\hat{g}_{h,h'} - \bar{g}_{h,h'}) + \bar{g}_{h'} - \bar{g}_{h,h'}.$$

Using again the Minkowski inequality and the elementary inequality $(x + y + z)^2 \leq 3(x^2 + y^2 + z^2)$, $(x, y, z) \in \mathbb{R}^3$, we obtain

$$\|\hat{g}_{h'} - \hat{g}_{h,h'}\|_2^2 \leq 3 (\|\hat{g}_{h'} - \bar{g}_{h'}\|_2^2 + \|\hat{g}_{h,h'} - \bar{g}_{h,h'}\|_2^2 + \|\bar{g}_{h'} - \bar{g}_{h,h'}\|_2^2).$$

Set $D = [-\pi/h', -\pi(1/h \wedge 1/h')] \cup [\pi(1/h \wedge 1/h'), \pi/h']$. The Parseval theorem yields

$$\|\bar{g}_{h'} - \bar{g}_{h,h'}\|_2^2 = \frac{1}{2\pi} \int_D |g^*(t)|^2 dt \leq \int_{|t| \geq \pi/h} |g^*(t)|^2 dt = \|\bar{g}_h - g\|_2^2.$$

In the same way, we prove that

$$\|\hat{g}_{h,h'} - \bar{g}_{h,h'}\|_2^2 \leq \|\hat{g}_{h'} - \bar{g}_{h'}\|_2^2.$$

Then

$$\|\hat{g}_{h'} - \hat{g}_{h,h'}\|_2^2 \leq 6\|\hat{g}_{h'} - \bar{g}_{h'}\|_2^2 + 3\|g - \bar{g}_h\|_2^2. \tag{5.18}$$

It follows from (5.17), (5.18) and the Parseval theorem that

$$\mathbb{E}(A(h)) \leq 6U + 3 \int_{|t| \geq \pi/h} |g^*(t)|^2 dt, \quad (5.19)$$

where

$$U = \mathbb{E} \left(\sup_{h' \in \mathcal{H}_n} (\|\hat{g}_{h'} - \bar{g}_{h'}\|_2^2 - V(h')/6)_+ \right).$$

Upper bound for U. Let us consider the random event

$$\Omega(u) = \left\{ |\tilde{f}^*(u) - f^*(u)| \leq \tau \sqrt{\log(n)/n} \right\}$$

Let us set

$$v_n = \frac{h'}{2\pi} V(h') = \kappa \frac{(\log(n))^m}{n^m}.$$

Owing to the Parseval theorem and Lemma 5.2, we have

$$\begin{aligned} \|\hat{g}_{h'} - \bar{g}_{h'}\|_2^2 - V(h')/6 &= \frac{1}{2\pi} \int_{-\pi/h'}^{\pi/h'} |(\tilde{f}^*(t))^m - (f^*(t))^m|^2 dt - V(h')/6 \\ &\leq \frac{1}{\pi} \left(\int_{-\pi/h'}^{\pi/h'} (|\tilde{f}^*(t) - f^*(t)|^{2m} \mathbf{1}_{\Omega(t)} - v_n/12) dt + \int_{-\pi/h'}^{\pi/h'} |\tilde{f}^*(t) - f^*(t)|^{2m} \mathbf{1}_{\Omega^c(t)} dt \right. \\ &\quad \left. + D_m^2 \int_{-\infty}^{\infty} |\tilde{f}^*(t) - f^*(t)|^2 |f^*(t)|^2 dt \right). \end{aligned}$$

Therefore

$$U \leq C(R + S + T), \quad (5.20)$$

where

$$\begin{aligned} R &= \sum_{k=1}^{n^m} \mathbb{E} \left(\left(\int_{-\pi/h_k}^{\pi/h_k} (|\tilde{f}^*(t) - f^*(t)|^{2m} \mathbf{1}_{\Omega(t)} - v_n/12) dt \right)_+ \right), \\ S &= \sum_{k=1}^{n^m} \int_{-\pi/h_k}^{\pi/h_k} \mathbb{E} (|\tilde{f}^*(t) - f^*(t)|^{2m} \mathbf{1}_{\Omega^c(t)}) dt \end{aligned}$$

and

$$T = \int_{-\infty}^{\infty} \mathbb{E} (|\tilde{f}^*(t) - f^*(t)|^2) |f^*(t)|^2 dt.$$

Evaluation of R. On $\Omega(t)$, we have

$$|\tilde{f}^*(t) - f^*(t)| \leq \tau \sqrt{\frac{\log(n)}{n}}.$$

Therefore, taking τ large enough, we have

$$\int_{-\pi/h_k}^{\pi/h_k} \left(|\tilde{f}^*(t) - f^*(t)|^{2m} \mathbf{1}_{\Omega(t)} - v_n/12 \right) dt \leq 0.$$

Hence

$$R = 0. \tag{5.21}$$

Upper bound for S. Following Neumann (1997), we apply Bernstein Inequality and get that

$$\begin{aligned} \mathbb{P}(\Omega(t)^c) &= \mathbb{P} \left(\left| \frac{1}{n} \sum_{k=1}^n (e^{-itX_k} - \mathbb{E}(e^{-itX_k})) \right| \geq \tau \sqrt{\log(n)/n} \right) \\ &\leq 2(\exp(-\tau^2 \log(n)/4) + \exp(-3\tau \sqrt{n \log(n)}/4)). \end{aligned}$$

Thus, taking $\tau \geq 2 + p$ implies

$$\mathbb{P}(\Omega^c(t)) \leq Cn^{-p}. \tag{5.22}$$

Applying the Cauchy-Schwarz inequality, Lemma 5.1 with $v = 2m$, (5.22) with $p = m + 2$ and $1/h_k \leq n^m$, we obtain

$$\begin{aligned} S &\leq \sum_{k=1}^{n^m} \int_{-\pi/h_k}^{\pi/h_k} \left(\mathbb{E}(|\tilde{f}^*(t) - f^*(t)|^{4m}) \right)^{1/2} (\mathbb{P}(\Omega^c(t)))^{1/2} dt \\ &\leq C \frac{1}{n^{m+p/2}} \sum_{k=1}^{n^m} \frac{1}{h_k} \leq C \frac{1}{n^{(p-m)/2}} = C \frac{1}{n}. \end{aligned} \tag{5.23}$$

Note that we get $\kappa \geq 4 + m$ for the constant in the definition of $V(h)$.

Upper bound for T. Using Lemma 5.1 with $v = 1$ and applying the Parseval theorem, we get

$$T \leq C \frac{1}{n} \|f^*\|_2^2 \leq C \frac{1}{n} \|f\|_2^2 \leq C \frac{1}{n}. \tag{5.24}$$

Putting (5.20), (5.21), (5.23) and (5.24) together, we have

$$U \leq C \frac{1}{n}. \tag{5.25}$$

Combining (5.19) and (5.25), we get

$$\mathbb{E}(A(h)) \leq C \left(\frac{1}{n} + \int_{|t| \geq \pi/h} |g^*(t)|^2 dt \right). \tag{5.26}$$

Finally, from (5.16) and (5.26), we obtain

$$\mathbb{E}(\|\hat{g}_h - g\|_2^2) \leq C \left(\frac{1}{hn^m} + \int_{|t| \geq \pi/h} |g^*(t)|^2 + V(h) + \frac{1}{n} \right). \quad (5.27)$$

We obtain the desired result by taking the infimum of h over \mathcal{H}_n .

Rates of convergence.

(vii) $m \geq 2$, $\beta > 1/(2m(m-1))$: It follows from (5.27) and (5.15) with $h = [n^{-m/(2m\beta+1)}]$ that

$$\mathbb{E}(\|\hat{g}_h - g\|_2^2) \leq C \left(n^{-2m^2\beta/(2m\beta+1)} + (\log(n))^m n^{-2m^2\beta/(2m\beta+1)} + \frac{1}{n} \right) \leq C \frac{1}{n}.$$

(viii) $\beta \in (0, 1/(2m(m-1))]$: Putting (5.27) and (5.15) together with

$$h = \left[\left(\frac{(\log(n))^m}{n} \right)^{m/(2m\beta+1)} \right]$$

we obtain

$$\mathbb{E}(\|\hat{g}_h - g\|_2^2) \leq C \left(\left(\frac{(\log(n))^m}{n} \right)^{2m^2\beta/(2m\beta+1)} + \frac{1}{n} \right) \leq C \left(\frac{(\log(n))^m}{n} \right)^{2m^2\beta/(2m\beta+1)}.$$

Theorem 3.1 is complete. □

References

- Ahmad, I.A. and Fan, Y. (2001). Optimal bandwidth for kernel density estimators of functions of observations. *Statist. Probab. Lett.*, 51, (3), 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.
- Caroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83, 1184-1186.
- Cherubini, U., Mulinacci, S. and Romagnoli, S. (2011) On the distribution of the (un)bounded sum of random variables. *Insurance Math. Econom.* 48, no. 1, 56-63.

- Comte, F. and Genon-Catalot, V. (2012). Convolution power kernels for density estimation. *Journal of Statistical Planning and Inference*, 142, 7, 1698-1715.
- 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.
- 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.
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39, 1608-1632.
- Jones, M.C. and Lotwick, H.W. (1984) Remark AS R50. A remark on algorithm AS 176. *Appl. Statist.*, 33 120-122.
- Lepski, O.V. (1990). On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35, 454-466.
- Marron, J.S. and Wand, M.P. (1992), Exact Mean Integrated Squared Error. *Ann. Statist.*, 20, 712-736.
- Mugdadi, A.R. and Ahmad, I. (2004). A Bandwidth Selection for Kernel Density Estimation of Functions of Random Variables. *Computational Statistics and Data Analysis*, 47,1, 49-62.
- Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.*, 7, 307-330.
- Neumann, M. and Reiss, M. (2009), Nonparametric estimation for Levy processes from low-frequency observations. *Bernoulli*, 15, 223-248.
- Nickl, R. (2007) Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields*, 138, no. 3-4, 411-449.
- Nickl, R. (2009) On convergence and convolutions of random signed measures. *J. Theoret. Probab.*, 22, 38-56.

- Panjer, H.H. and Willmot, G.E. (1992). Insurance Risk Models. Society of Actuaries, Schaumburg.
- 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.
- Rosenthal, H.P. (1970). On the subspaces of \mathbb{L}_p ($p \geq 2$) spanned by sequences of independent random variables. *Israel Journal of Mathematics*, 8, 273-303.
- Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, 28, 799-815.
- 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.
- Silverman B.W. (1982), Algorithm AS 176. Kernel density estimation using the fast Fourier transform. *Appl. Statist.*, 31, 93-99.
- Silverman, B. W. (1986), Density estimation: for statistics and data analysis, Chapman & Hall.
- Tsybakov, A.B. (2004). *Introduction à l'estimation non-paramétrique*, Springer.
- Zhang, C.-H. (2005). Estimation of sums of random variables: examples and information bounds. *Ann. Statist.*, 33, no. 5, 2022-2041.
- Zijaeva, Z. T. (1975) The estimation of the density of a sum of random variables. (Russian) *Izv. Akad. Nauk. USSR Ser. Fiz.-Mat. Nauk*, no. 3, 13-15, 95.