Density estimation for nonnegative random variables

Fabienne Comte, Valentine Genon-Catalot

DENSITY ESTIMATION FOR NONNEGATIVE RANDOM VARIABLES

F. COMTE[(1)*] AND V. GENON-CATALOT[(1)]

Abstract. We propose a new type of non parametric density estimators fitted to nonnegative random variables. The estimators are constructed using kernels which are densities of empirical means of m i.i.d. nonnegative random variables with expectation 1. The value $m^{-1/2}$ plays the role of the bandwidth. We study the pointwise Mean Square Error and a weighted global Mean Integrated Square Error and propose adaptive estimators for both local and global points of view. The risks of the adaptive estimators satisfy oracle inequalities. A noteworthy result is that the adaptive rates are in correspondence with the smoothness properties of the unknown density as a function on $(0, +\infty)$. Pointwise adaptive esimators are illustrated on simulated data.

Keywords. Adaptive estimators. Density estimation. Infinitely divisible distributions. Kernel estimators. Nonnegative random variables.

1. Introduction

Let $X_1, \ldots, X_n$ be n *i.i.d.* nonnegative random variables, with unknown density f. The problem of estimating f as a function on $\mathbb{R}$ by a non-parametric approach has received a lot of attention in the past decades. A huge variety of methods have been investigated and powerful techniques allow to build concrete estimators having optimality properties in the sense that their $\mathbb{L}^2$-risk reaches automatically the best possible rate associated with the unknown smoothness of the unknown density. Among many authors, we can quote Stone (1980), Devroye and Lugosi (1996), Donoho *et al.* (1996), Kerkyacharian *et al.* (1996), Birgé and Massart (1997), Juditsky (1997), Barron *et al.* (1999). See also Butucea (2001) for a detailed discussion.

In the case of nonnegative observations, $f = 0$ on $(-\infty, 0)$. Hence, even if f is very smooth as a function on $(0, +\infty)$, f may be not even continuous as a function on $(-\infty, +\infty)$. Therefore, it is not wise to use blindly estimators fitted with functions on the whole real line. One should at least use a transformation of the data (such as taking logarithms) and estimate the density of the transformed variables. Thus, there is a need to find specific methods to estimate densities for nonnegative or lower bounded random variables. Nonnegative random variables are commonly used for models in survival analysis or reliability, for waiting times in renewal processes ...

In this paper, our aim is to propose a new type of estimators fitted to nonnegative random variables which can be used directly without transforming data. The basic idea of their construction is the following. Consider a density K on $(0, +\infty)$, with expectation 1, and let $U_1, \ldots, U_m$ be m *i.i.d.* random variables with distribution $K(u)du$ ($K = 0$ on $(-\infty, 0)$). The empirical mean $\bar{U}_m = (U_1 + \ldots + U_m)/m$ has distribution $K_m(u)du$ with

$$K_m(u) = mK^{\star m}(mu) \tag{1}$$

[(1)] Université Paris Descartes, MAP5, UMR CNRS 8145.

[*] Corresponding author: Fabienne Comte, MAP5 UMR 8145, Université Paris Descartes, 45 rue des Saints Pères, 75270 PARIS cedex 06, FRANCE. email: fabienne.comte@parisdescartes.fr.

where $K^{\star m} = K \star \cdots \star K$, m times and $\star$ denotes the convolution product. The distribution $K_m(u)du$ converges weakly to 1 as m grows to infinity. For given $x > 0$, we propose to estimate the value $f(x)$ by

$$\hat{f}_m(x) = \frac{1}{nx} \sum_{k=1}^{n} K_m(\frac{X_k}{x}). \tag{2}$$

The noteworthy property of the above estimator is that, by an elementary change of variable,

$$\mathbb{E}\hat{f}_m(x) = \mathbb{E}f(x\bar{U}_m),$$

which is close to $f(x)$ for large m *e.g.* when f is continuous at x and bounded. If f is smoother, we can use a Taylor expansion of the difference $f(x\bar{U}_m) - f(x)$ and integrate it. Note that, by the assumption on K, the first term $(\bar{U}_m - 1)xf'(x)$ is centered. In the standard case where K has finite variance v, the second term in the Taylor expansion being $(x^2/2)(\bar{U}_m - 1)^2 f''(x)$ has expectation of order m^{-1}. Below, we prove that $\mathrm{Var}\hat{f}_m(x)$ is of order $\sqrt{m}/n$. Choosing m such that $m^{-2} = \sqrt{m}/n$ (the usual square bias-variance compromise) yields the standard rate $n^{-4/5}$ for the estimation of a C^2 function at x.

These preliminary remarks motivate a more thorough study and raise the following question: is it possible to prove that estimators such as (2) to estimate functions on $\mathbb{R}^+$ fulfill all properties of non-parametric kernel estimators of functions on $\mathbb{R}$. This is the topic of the present paper and the answer is yes. First, we extend the definition of the estimator to obtain the best possible rate for the bias term fitted with the smoothness of the unknown density (as a function on $(0, +\infty)$). For this, we define a new notion of kernel of order ℓ and study the pointwise estimation. Then, we study the global properties by introducing a weighted Mean Integrated squared Error (MISE). The value $m^{-1/2}$ plays the role of the bandwidth and we study in both cases (pointwise and global) adaptive bandwidth selection. For this, we prove that the method of Goldenshluger and Lepski (2010) can be adapted to our estimators.

In Section 2, we precise the assumptions on the densities K used to build estimators, extend the definition of (2) and prove a key preliminary result (Lemma 2.1) on the Fourier transform of K_m, which enables us to obtain the variance rate of our estimators. We give examples of densities K for which K_m is explicit. The case of K an infinitely divisible density is especially interesting, as it allows to give an explicit construction of kernels of order ℓ. It also allows to extend the definition of K_m to the case of non integer m. This construction is closely related to the construction of Kerkyacharian *et al.* (2001). Section 3 is devoted to pointwise estimation. We study the pointwise Mean Square Error and build an adaptive estimator fitted to local Hölder regularity. For adaptive bandwidth selection, we introduce iterated kernels built using densities of the form $K_m \odot K_{m'}$ where $K_1 \odot K_2$ denotes the density of the product UV where U, V are independent nonnegative variables, U has density K_1 and V has density K_2. We prove in Theorem 3.1 that the estimator reaches the same optimal adaptive rate as the one obtained by Butucea (2001) on Sobolev classes. In Section 4, we study a global weighted Mean Integrated Square Error (MISE). This requires the introduction of a new type of Nikol'skii class of functions on $(0, +\infty)$. Classical examples of densities are shown to belong to this class. We propose a global adaptive estimator still based on iterated kernels. Theorem 4.2 gives an oracle inequality for the weighted MISE of this estimator. In Section 5, numerical simulation results are provided for local adaptive selection of m. Moreover, we consider the kernel estimators described in Goldenschluger and Lepski (2010b) and implement a local adaptive bandwidth selection which is new. Both methods are convincing and we discuss their respective advantages and drawbacks. Section 6 contains concluding remarks. Proofs are gathered in Section 7.

2. Definition of estimators and kernels.

2.1. **Assumptions and preliminaries.** For K a density on $\mathbb{R}^+$, we consider the following assumptions:

- (H1) For $u \geq 0$, $K(u) \geq 0$, for $u < 0$, $K(u) = 0$ and

$$\int_0^{+\infty} K(u)du = 1, \int_0^{+\infty} uK(u)du = 1, \int_0^{+\infty} (u-1)^2 K(u)du = v, \int_0^{+\infty} K^2(u)du < +\infty.$$

- (H2) Let K_m be defined by (1). For m large enough, $I_m := \int K_m(u)\frac{du}{u} = 1 + O(\frac{1}{m})$.
- (H3) Let $\nu_\gamma = \int |u-1|^\gamma K(u)du$. There exists $\gamma \geq 4$ such that $\nu_\gamma < +\infty$.

A density K satisfying (H1) has expectation equal to 1, finite variance $v(K) = v$ and belongs to $\mathbb{L}^2$. The function K_m defined by (1) is the density of the empirical mean $\bar{U}_m = (U_1 + \ldots + U_m)/m$ of m *i.i.d.* random variables $U_1, \ldots, U_m$ with distribution $K(u)du$. The Fourier transform of K:

$$K^*(t) = \int_0^{+\infty} e^{itu} K(u)du$$

belongs to $\mathbb{L}^2(\mathbb{R})$. The Fourier transform of K_m is $K_m^*(t) = K^*(t/m)^m$. As K^* is a characteristic function, $|K^*(t)| \leq 1$ for all $t \in \mathbb{R}$. Thus, for all $m \geq 2$, $K_m^*(t)$ belongs to $\mathbb{L}^1(\mathbb{R})$ and for all $m \geq 1$, $K_m^*(t)$ belongs to $\mathbb{L}^2(\mathbb{R})$. The rate of the variance term in the risk of our estimator relies on the following lemma.

Lemma 2.1.

- *Let K be a density satisfying* (H1). *For $\alpha \geq 1$, we have*

$$(\sqrt{m})^{-1} \int_{\mathbb{R}} |K_m^*(t)|^\alpha dt \to_{m \to +\infty} \sqrt{2\pi/(\alpha v)}.$$

Consequently, as m tends to infinity,

$$||K_m||_\infty \leq \sqrt{m}(1/\sqrt{2\pi v})(1 + o(1)) \text{ and } ||K_m||_2^2 = \sqrt{m}(1/2\sqrt{\pi v})(1 + o(1)),$$

where $||K_m||_\infty = \sup_{u \geq 0} K_m(u)$ and $||K_m||_2^2 = \int_0^{+\infty} K_m^2(u)du$.

- *Let $K^{(1)}, K^{(2)}$ be two densities satisfying* (H1) *with variances v_1, v_2 respectively. As m tends to infinity,*

$$< K_m^{(1)}, K_m^{(2)} >= \int_0^{+\infty} K_m^{(1)}(u) K_m^{(2)}(u)du = \sqrt{m}(1/\sqrt{2\pi(v_1 + v_2)})(1 + o(1)).$$

Integrating formula (2) on $\mathbb{R}^+$, we get:

$$\int_0^{+\infty} \hat{f}_m(x)dx = \int_0^{+\infty} K_m(u)\frac{du}{u} = \mathbb{E}(\frac{1}{\bar{U}_m}) := I_m.$$

Thus, we have to impose that this expectation be finite at least for m large enough (Assumption (H2)). This is not a problem as we choose the kernel. In such a case, as $\mathbb{E}\bar{U}_m = 1$, we have

$$\mathbb{E}(\frac{1}{\bar{U}_m}) - 1 = \frac{1}{m}\,\mathbb{E}(m\frac{(1-\bar{U}_m)^2}{\bar{U}_m})$$

Knowing that $m\frac{(1-\bar{U}_m)^2}{\bar{U}_m}$ converges in distribution to vZ^2 with Z a standard Gaussian variable, we can easily choose the kernel K such that $\mathbb{E}(m\frac{(1-\bar{U}_m)^2}{\bar{U}_m}) = O(1)$, so that $I_m = 1 + O(1/m)$ (see below the examples). Thus, $I_m^{-1}\hat{f}_m(x)$ is a density.
The moment assumption (H3) is used to evaluate the rate of the absolute moments

$$\nu_{\gamma,m} = \int |u-1|^\gamma K_m(u)du,$$

(see below Lemma 7.1).

Remark 2.1. *Note that we can link formula (2) with a more usual kernel type estimator as follows. Let $k_m(z)$ denote the density of $\sqrt{m/v}(\bar{U}_m - 1)$. The following relations holds:*

$$k_m(z) = \frac{\sqrt{v}}{\sqrt{m}} K_m(1 + z\frac{\sqrt{v}}{\sqrt{m}}),$$

and setting $h_m(x) = \sqrt{v}x/\sqrt{m}$,$\hat{f}_m(x) = (nh_m(x))^{-1} \sum_{k=1}^n k_m\left((X_i - x)/h_m(x)\right)$. Hence, $\hat{f}_m(x)$ looks like a kernel estimator with kernel k_m and bandwidth $h_m(x)$.

2.2. **Extension of the definition of estimators.** In order to have adequate properties of the bias term, we must consider a more general class of estimators than the one defined by (2) and introduce kernels that can be real-valued. Let $\alpha_1, \ldots, \alpha_L$ be real numbers such that

$$\sum_{j=1}^{L} \alpha_j = 1 \tag{3}$$

and consider the extended kernel $K = \sum_{j=1}^L \alpha_j K^{(j)}$ and set, for $m \geq 1$,

$$K_m = \sum_{j=1}^{L} \alpha_j K_m^{(j)} \tag{4}$$

where $K^{(j)}$ is a density on $\mathbb{R}^+$ satisfying assumptions (H1)-(H3) with variance denoted by v_j, and $K_m^{(j)}(u) = m(K^{(j)})^{\star m}(mu)$ is the density of $(U_1^j + \ldots U_m^j)/m$ with $U_k^j, k = 1, \ldots, m$ i.i.d. with density $K^{(j)}$ (for $m = 1$, $K_1 = K, K_1^{(j)} = K^{(j)}$). The functions K_m are not necessarily positive. We define

$$\hat{f}_m(x) = \sum_{j=1}^{L} \alpha_j \hat{f}_{m,j}(x) = \frac{1}{nx} \sum_{k=1}^{n} K_m(\frac{X_k}{x}). \tag{5}$$

where $\hat{f}_{m,j}(x)$ is given by formula (2) using the density $K_m^{(j)}$. By Assumption (H2) and (3), $\int_0^{+\infty} \hat{f}_m(x)dx = 1 + O(1/m)$.

2.3. **Examples of kernels yielding explicit formulae.** The method proposed here leads to explicit formulae when the m-th convolution of a density K satisfying (H1)-(H3) can be explicitly computed. We can note below that m need not be integer.

2.3.1. *Gamma kernels.* For K the Gamma density $G(a, a)$ for any positive a, (H1)-(H3) hold and:

$$K_m(u) = \frac{(am)^{am}}{\Gamma(am)} u^{am-1} e^{-mau} 1_{u>0}.$$

For $m > 1/a$, $I_m = \int_0^{+\infty} u^{-1} K_m(u)du = (am)/(am - 1) = 1 + O(1/m)$.

In particular, for $K(u) = e^{-u}1_{u>0}$ (exponential density with parameter 1), the estimator (2) is given by

$$\hat{f}_m(x) = \frac{1}{n}\left(\frac{m}{x}\right)^m \frac{1}{(m-1)!} \sum_{k=1}^{n} X_k^{m-1} e^{-\frac{m}{x}X_k}. \tag{6}$$

Introducing the Laplace transform of X, $L_X(\lambda) = \int_0^{+\infty} e^{-\lambda t} f(t)dt$, for $\lambda \geq 0$, and its successive derivatives $L_X^{(j)}(\lambda) = (-1)^j \int_0^{+\infty} t^j e^{-\lambda t} f(t)dt$, we have, for $x > 0$:

$$\mathbb{E}\hat{f}_m(x) = \Big(\frac{m}{x}\Big)^m \frac{(-1)^{m-1}}{(m-1)!} L_X^{(m-1)}(\frac{m}{x}). \tag{7}$$

This is a very well known inversion formula for Laplace transforms to get an approximation of $f(x)$ (see *e.g.* Feller, 1971).

2.3.2. *Inverse Gaussian kernels.* The inverse Gaussian distribution $IG(a, \theta)$ $a > 0, \theta > 0$, is defined as the distribution of the hitting time $T_a = \inf\{t \geq 0,\ \theta t + B_t = a\}$ where (B_t) is a Wiener process. This parametrization is the one used in Barndorff-Nielsen (1998). The density of an $IG(a, \theta)$ is

$$\frac{a}{\sqrt{2\pi t^3}} e^{\theta a} e^{-\frac{1}{2}\left(\theta^2 t + \frac{a^2}{t}\right)}.$$

Its Laplace transform is given by

$$\mathbb{E}(e^{-\lambda T_a}) = \exp\left(a(\theta - \sqrt{\theta^2 + 2\lambda})\right),\ \lambda > -\theta^2/2.$$

Thus for $a = \theta$, the expectation is 1 and the variance is $v = 1/a^2$.

For K the inverse Gaussian density $IG(a, a)$, (H1)-(H3) hold and K_m is the density of the law $IG(a\sqrt{m}, a\sqrt{m})$:

$$K_m(u) = \frac{a\sqrt{m}}{\sqrt{2\pi u^3}} e^{ma^2(1-\frac{1}{2}(\frac{1}{u}+u))} 1_{u>0}.$$

Using properties of the Generalized Inverse distributions (see Barndorff-Nielsen and Shephard (2001), p.173), we can compute:

$$I_m = \int_0^{+\infty} u^{-1} K_m(u)du = \frac{a\sqrt{m}}{\sqrt{2\pi}} e^{ma^2} 2\tilde{K}_{-3/2}(a^2 m),$$

where $\tilde{K}_\nu$ is the modified Bessel function of second kind (see Abramowitz and Stegun (1964), 9.6.23 p.376)

$$\tilde{K}_\nu(u) = \tilde{K}_{-\nu}(u) = \frac{\sqrt{\pi}(z/2)^\nu}{\Gamma(\nu + 1/2)} \int_0^{+\infty} e^{-z\cosh(t)} (\sinh t)^{2\nu} dt. \tag{8}$$

We use the notation $\tilde{K}_\nu$ for the Bessel function to avoid confusion with the density K_m.

After the successive changes of variables $e^t = u$ and $s = u + u^{-1} - 2$, we obtain

$$e^z \tilde{K}_{3/2}(z) = \sqrt{\frac{\pi}{2}}(\frac{1}{\sqrt{z}} + \frac{1}{z^{3/2}}).$$

We thus find $I_m = 1 + 1/(a^2 m)$.

3. Pointwise Estimation.

In this section, we study the properties of (5) as an estimator of the value $f(x)$ for a given positive x.

3.1. **Pointwise MSE.** The pointwise Mean Squared Error (MSE) is given by $\mathbb{E}(\hat{f}_m(x)-f(x))^2$ and is split into the usual sum of the variance and squared bias terms.

First, we give the variance bound.

Proposition 3.1. *Assume that f is bounded and let $\hat{f}_m(x)$ be defined by (5) with (3)-(4) . Then, for all n, as m tends to infinity,*

$$Var\hat{f}_m(x) \leq \frac{\sqrt{m}||f||_\infty}{nx}\frac{1}{\sqrt{2\pi}}\left(\sum_{1\leq i,j\leq L}\frac{\alpha_i\alpha_j}{\sqrt{v_i+v_j}}\right)(1+o(1)),$$

where $o(1)$ is uniform in x (v_j denotes the variance of the density $K^{(j)}$). If $L=1$ the constant in parenthesis is just $1/\sqrt{2v}$.

To study the bias term, the notion of kernel of order ℓ can be here defined as follows.

Definition 3.1. *We say that $K=\sum_{j=1}^L \alpha_j K^{(j)}$ defines a kernel of order ℓ if, for $j=1,\ldots,L$, the kernel $K^{(j)}$ satisfies Assumptions* (H1)-(H3) *(with variance v_j), admits moments up to order ℓ and the coefficients $\alpha_j, j=1,\ldots,L$ are such that $\sum_{j=1}^L \alpha_j = 1$ and for $1\leq k\leq \ell$ and all m (at least large enough)*

$$\int_0^{+\infty}(u-1)^k K_m(u)du = \sum_{j=1}^L \alpha_j \int_0^{+\infty}(u-1)^k K_m^{(j)}(u)du = 0.$$

These relations allow to compute the α_j's as functions of the moments of the $K^{(j)}$'s. Below, we show how to compute kernels of order ℓ in the above sense.

The Hölder-type class on $(0,\infty)$ with constant C and regularity β is defined in its usual sense:

$$\begin{aligned}\Sigma(\beta,C) \;=\; &\{f:(0,+\infty)\to\mathbb{R},\\ &f^{(\ell)}\text{exists for } \ell=\lfloor\beta\rfloor \text{and } |f^{(\ell)}(x)-f^{(\ell)}(x')|\leq C|x-x'|^{\beta-\ell}, \text{for all } x,x'>0\}.\end{aligned}$$

We obtain the following bias bound.

Proposition 3.2. *Assume that f belongs to the class $\Sigma(\beta,C)$. Assume that, for $j=1,\ldots,L$, $\nu_\beta^{(j)}=\int|u-1|^\beta K^{(j)}(u)du<+\infty$ and that K is a kernel of order $\ell=\lfloor\beta\rfloor$ in the sense of 3.1. Then,*

$$|\mathbb{E}\hat{f}_m(x)-f(x)|\leq \frac{C\;C(\beta)x^\beta}{\ell!}\frac{1}{m^{\beta/2}}$$

where $C(\beta)$ depends on the coefficients α_j and the moments of the $K^{(j)}$'s.

Remark 3.1. *The assumption that the $K^{(j)}$'s are square integrable is not required in this result.*

Finally, for the pointwise quadratic risk, we can state:

Proposition 3.3. *Assume that f is bounded and belongs to the class $\Sigma(\beta,C)$. Assume that, for $j=1,\ldots,L$, $\nu_\beta^{(j)}=\int|u-1|^\beta K^{(j)}(u)du<+\infty$ and that K is a kernel of order $\ell=\lfloor\beta\rfloor$ in the sense of 3.1. For all n and all positive x, as m tends to infinity,*

$$\mathbb{E}\left(\hat{f}_m(x)-f(x)\right)^2\leq\frac{C_1\sqrt{m}}{nx}+(C_2(\beta))^2\frac{x^{2\beta}}{m^\beta},$$

where

$$C_1=\frac{(1+o_m(1))}{\sqrt{2\pi}}||f||_\infty\left(\sum_{1\leq i,j\leq L}\frac{\alpha_i\alpha_j}{\sqrt{v_i+v_j}}\right),\quad C_2(\beta)=C\frac{C(\beta)}{\lfloor\beta\rfloor!},\tag{9}$$

where $o_m(1)$ does not depend on x and $C(\beta)$ is the constant appearing in the bias bound in Proposition 3.2.

Note that

$$\sum_{1\le i,j\le L} \frac{\alpha_i\alpha_j}{\sqrt{v_i+v_j}} \le \sum_{i=1}^{L} \frac{|\alpha_i|}{\sqrt{v_i}} \sum_{i=1}^{L} |\alpha_i|.$$

The choice of m leading to the optimal bias-variance compromise is obtained for $x^{2\beta}m^{-\beta} = (nx)^{-1}m^{1/2}$, *i.e.* $m = x^2 n^{2/(2\beta+1)}$ and the rate of the pointwise risk is $n^{-2\beta/(2\beta+1)}$ and does not depend on x. This rate is the standard optimal rate for estimating a density in the class $\Sigma(\beta, C)$ (see Tsybakov (2009) and the references therein).

3.2. **Kernels of order ℓ.** In this paragraph, we show how to build a kernel of order ℓ as defined in 3.1. Let K be a density on $\mathbb{R}^+$ satisfying (H1) and which is infinitely divisible with Lévy density $n(.)$ on $\mathbb{R}^+$. The characteristic function of K is given by:

$$K^*(t) = \exp\left(\int_0^{+\infty} \left(e^{itx}-1\right) n(x)dx\right) = \exp\left(it + \psi(t)\right), \tag{10}$$

with

$$\psi(t) = \int_0^{+\infty} \left(e^{itx}-1-itx\right) n(x)dx,$$

see *e.g.* Sato (1999). By the assumption on K, we have, by twice derivating K^*,

$$1 = \int_0^{+\infty} uK(u)du = \int_0^{+\infty} xn(x)dx, \quad v = \int_0^{+\infty} (u-1)^2K(u)du = \int_0^{+\infty} x^2n(x)dx.$$

The distribution with density K has moment up to order k if and only if $\int_0^{+\infty} x^k n(x)dx < +\infty$. Moreover, the j-th order cumulant of K obtained by developping $it+\psi(t)$ around 0 is simply given using the Lévy density by

$$\kappa_j = \int_0^{+\infty} x^j n(x)dx.$$

Proposition 3.4. *Assume that K is a density satisfying* (H1)*, admitting moments up to order n, with characteristic function given by (10). For $a > 0$, define the distribution $P_a(du)$ on $\mathbb{R}^+$ by the characteristic function*

$$K_a^*(t) = \left(K^*(\frac{t}{a})\right)^a = e^{it+\psi_a(t)}, \tag{11}$$

where

$$\psi_a(t) = \int_0^{+\infty} \left(e^{itx}-1-itx\right) n_a(x)dx, \quad \text{with } n_a(x) = a^2n(ax).$$

For $a > 2$, the distribution P_a has a density K_a satisfying (H1) *and admitting moments up to order n. Let $\mu_{j,a} = \mathbb{E}(V_a^j)$ be the j-th moment of $V_a = U_a - 1$ where U_a has density K_a. Then,*

$$\mu_{n,a} = \sum_{\ell=k(n)}^{n-1} \frac{c(\ell)}{a^\ell}$$

where $c(n-1) = \kappa_n$ is the n-th cumulant of K, the other constants $c(\ell)$ depend on the cumulants $\kappa_\ell, 2 \le \ell \le n-1$ of K and $k(n) = n/2$ for n even, $k(n) = (n+1)/2$ for n odd.

It is worth noting that, for all integer m,

$$K^*_{am}(t) = \left(K^*(\frac{t}{am})\right)^{am}$$

is the characteristic function of the density $mK_a^{\star m}(u)$.

Remark 3.2.
- *Using the density K_a, we can extend the definition of the estimator by changing K_m into K_a for any positive a by setting:*

$$\hat{f}_a(x) = \frac{1}{nx}\sum_{k=1}^{n} K_a(\frac{X_k}{x}) \tag{12}$$

If U_a has distribution $K_a(u)du$, $U_a \to 1$ and $\sqrt{a}(U_a - 1)$ tends in distribution to $\mathcal{N}(0, v)$.
- *The examples given above fit in this framework. The density Gamma $G(a_0, a_0)$ has Lévy density $n(x) = x^{-1}a_0e^{-a_0x}1_{x>0}$. The Inverse Gaussian density $IG(a_0, a_0)$ has Lévy density $n(x) = a_0(2\pi x^3)^{-1/2}e^{-a_0^2x/2}1_{x>0}$.*

We use the previous moments properties to build a kernel of order ℓ. The construction of Kerkyacharian *et al.* (2001) can be adapted to our kernels as follows.

Proposition 3.5. *Let K be a density satisfying* (H1)-(H3)*, admitting moments up to order $\ell \geq 2$, with characteristic function of the form (10). Let $a > 2\ell$ and consider the density K_a specified by (11). For $1 \leq j \leq \ell$ integer, set*

$$K^{(j)} = K_{aj^{-1}}$$

(see (11)) so that $K_m^{(j)} = K_{amj^{-1}}$. The function

$$K = \sum_{j=1}^{\ell} \binom{\ell}{j} (-1)^{j+1} K^{(j)}$$

is a kernel of order ℓ in the sense of definition 3.1.

3.3. **Local bandwidth selection.** We assume that f belongs to the class $\Sigma(\beta_0, C)$ where the true value β_0 is unknown, and $\beta_0 \leq \beta_{max}$ with known β_{max}. We consider kernels $K = \sum_{i=1}^{L} \alpha_i K^{(i)}$ of order $l \geq \beta_{max}$. Recall the notations (9) and set

$$B = \sup_{\beta \leq \beta_{max}} C_2(\beta).$$

The function $\beta \to C(\beta)$ is upper bounded on $[0, \beta_{max}]$, see also Proposition 7.1.

For $x_0 > 0$ and $m = m_{opt}(x_0) = x_0^2 n^{2/(2\beta_0+1)}$, the estimator $\hat{f}_{m_{opt}(x_0)}(x_0)$ satisfies:

$$\mathbb{E}\left(\hat{f}_{m_{opt}(x_0)}(x_0) - f(x_0)\right)^2 \leq (C_1 + B^2)n^{-2\beta_0/(2\beta_0+1)}.$$

Our aim is to define a data driven choice of m for the estimation of $f(x_0)$, leading to an estimator whose rate is as near as possible of the optimal one. This requires that $x_0 \geq x_{min} > 0$.

For two functions s and t on $(0, +\infty)$, let us define, each time it exists, the function defined on $(0, +\infty)$ by

$$u \to s \odot t(u) = \int_0^{+\infty} s(u/v)t(v)dv/v.$$

Consider U_1, U_2 two nonnegative independent random variables with densities K_1, K_2 respectively. Then, the density of the product U_1U_2 is equal to $K_1 \odot K_2(u)$.

In view of $m_{opt}(x_0)$, we could select m among values of the form $m = k^2 x_0^2$. This is possible from the theoretical point of view. In practice, the drawback is that the proposed choices for m can be very small if x_0 is small, which yields numerical problems. This is why we set

$$\mathcal{M}_n = \{m = k^2, \ \log(n) \le k \le n/\log(n)\} \tag{13}$$

as the set of possible indexes m and consider $K_m = \sum_{j=1}^{L} K_m^{(j)}$ an extended kernel as defined in (3)-(4) where the densities $K^{(j)}$ satisfy (H1)-(H3). Set

$$\hat{f}_{m,m'}(x) = \frac{1}{nx} \sum_{k=1}^{n} K_{m'} \odot K_m(\frac{X_k}{x}) = K_{m'} \odot \hat{f}_m(x). \tag{14}$$

As obviously $K_m \odot K_{m'} = K_{m'} \odot K_m$, we have $\hat{f}_{m,m'}(x) = \hat{f}_{m',m}(x)$. Note that

$$\hat{f}_{m,m'}(x) = \sum_{i,j=1}^{L} \alpha_i \alpha_j \frac{1}{nx} \sum_{k=1}^{n} K_m^{(i)} \odot K_{m'}^{(j)}(\frac{X_k}{x}). \tag{15}$$

For m, m' integers, $K_m^{(i)} \odot K_{m'}^{(j)}$ is the density of the product $(U_1^{(i)} + \ldots + U_m^{(i)})/m \times (V_1^{(j)} + \ldots + V_{m'}^{(j)})/m'$ where $U_1^{(i)}, \ldots, U_m^{(i)}, V_1^{(j)}, \ldots, V_{m'}^{(j)}$ are independent random variables, the $U_\ell^{(i)}$'s have density $K^{(i)}$ and the $V_\ell^{(j)}$'s have density $K^{(j)}$. So, as m, m' tend to infinity, $K_m^{(i)} \odot K_{m'}^{(j)}(u)du$ tends to 1.

Now, define

$$V(m, x_0) = \kappa C(K) \max(\|f\|_\infty, 1) \log(n) \frac{\sqrt{m}}{nx_0} \tag{16}$$

where κ is a numerical constant and

$$C(K) = 2(\sum_{i=1}^{L} |\alpha_i|)^3 (\sum_{i=1}^{L} |\alpha_i|/\sqrt{2\pi v_i}). \tag{17}$$

Note that, as $\sum_{i=1}^{L} \alpha_i = 1$, $\sum_{i=1}^{L} |\alpha_i| \ge 1$. In $C(K)$, the value 2 replaces $(1 + o_m(1))$ in formula (9). This is justified by the fact that we choose $m \ge \log^2(n)$, n large enough.

The following quantity must be understood as an approximation of the squared bias term:

$$A(m, x_0) = \sup_{m' \in \mathcal{M}_n} \left((\hat{f}_{m'}(x_0) - \hat{f}_{m,m'}(x_0))^2 - V(m', x_0) \right)_+ . \tag{18}$$

The adaptive estimator is then

$$\tilde{f}(x_0) = \hat{f}_{\hat{m}(x_0)}(x_0) \ \text{ with } \ \hat{m}(x_0) = \arg \min_{m \in \mathcal{M}_n} \left(A(m, x_0) + V(m, x_0) \right). \tag{19}$$

We can prove the following result.

Theorem 3.1. *Assume that f belongs to the class $\Sigma(\beta_0, C)$ with β_0 unknown, $0 < \beta_0 \le \beta_{max}$, with known β_{max}. Let K be a kernel of order ℓ in the sense of Definition 3.1 with $\ell \ge \beta_{max}$. Consider the estimator $\tilde{f}(x_0) = \hat{f}_{\hat{m}(x_0)}(x_0)$ with $x_0 \ge x_{min} > 0$. Then, there exists a constant κ in (16) such that, for n large enough,*

$$\mathbb{E}[(\tilde{f}(x_0) - f(x_0))^2] \lesssim \left(\frac{n}{\log n} \right)^{-2\beta_0/(2\beta_0+1)},$$

where $x_n \lesssim y_n$ means $x_n \le C y_n$ for some constant C.

As we can see, the rate is optimal up to a $\log(n)$ factor, but this is expected (see Butucea (2001)). The method is illustrated in Section 5. We also rewrite a local version of the method of Goldenschluger and Lepski (2010), and compare both methods.

3.4. **Iterated kernels.** The estimators $\hat{f}_m$ can easily be implemented for instance when choosing the densities $K^{(i)}$ among Gamma densities $G(a_i, a_i)$ or among Inverse Gaussian densities $IG(a_i, a_i)$. As seen above, it is important to note that the value m need not be an integer as the densities $K_m^{(i)}$, in the two previous cases, are respectively $G(a_i m, a_i m)$ and $IG(a_i\sqrt{m}, a_i\sqrt{m})$. In particular, when dealing with local bandwidth selection at x_0, the value of m may be proportional to x_0^2.

For implementing the adaptive selection method, one must compute $K_m^{(i)} \odot K_{m'}^{(j)}$. In case of $K^{(i)} = G(a_i, a_i)$, we must compute the density of a product of independent Gamma variables. Some relatively explicit formulae can be obtained, involving Bessel functions of high index. The case of $K^{(i)} = IG(a_i, a_i)$ is simpler.

Proposition 3.6. *Let $K_m^{(i)} = IG(a_i\sqrt{m}, a_i\sqrt{m})$. We have:*

$$K_m^{(i)} \odot K_{m'}^{(j)}(u) = \frac{a_i a_j \sqrt{mm'}}{\pi u^{3/2}} \exp\left(ma_i^2 + m'a_j^2\right)\tilde{K}_0(c_{ij}(u)),$$

where $\tilde{K}_0$ is the modified Bessel function of second kind with index 0 and

$$c_{ij}(u) = \left(a_i^4 m^2 + a_j^4 (m')^2 + a_i^2 a_j^2 mm'(u + \frac{1}{u})\right)^{1/2}$$

Proof of Proposition 3.6 Recall that

$$K_m^{(i)} \odot K_{m'}^{(j)}(u) = \int_0^{+\infty} K_m^{(i)}(\frac{u}{v}) K_{m'}^{(j)}(v) \frac{dv}{v}$$

Thus, we must compute

$$I_{ij}(u) = \int_0^{+\infty} \exp\left[-\frac{1}{2}(\frac{1}{v}\delta_{ij}^2 + v\gamma_{ij}^2)\right]\frac{dv}{v},$$

where

$$\delta_{ij}^2 = a_i^2 mu + + a_j^2 m', \quad \gamma_{ij}^2 = a_i^2 \frac{m}{u} + + a_j^2 m'.$$

The integral $I_{ij}(u)$ can be computed using the norming constant of Generalized Inverse Gaussian densities (see Barndorff-Nielsen and Shephard (2001)):

$$I_{ij}(u) = 2\tilde{K}_0(\delta_{ij}\gamma_{ij}),$$

where $\tilde{K}_0$ is the modified Bessel function of second kind with index 0. This gives the result.□
The Bessel function $\tilde{K}_0$ is available in Matlab. Various expressions are given in Abramowitz and Stegun (1964).

4. Weighted MISE.

We study a weighted mean integrated squared error with weight function $w(x) = x \wedge 1$ as we need the weight $w_1(x) = x$ for the variance term and the weight $w_2(x) = 1$ for the bias term.

4.1. **Risk bound.**

Proposition 4.1. *(Variance bound) Assume that* $K_m = \sum_{i=1}^L \alpha_i K_m^{(i)}$ *where* $\sum_{i=1}^L \alpha_i = 1$*, for* $i = 1, \dots, L$*, the density* $K^{(i)}$ *satifies Assumptions* (H1) *and* (H2) *(with variance denoted by* v_i*). Then,*

$$\int_0^{+\infty} Var\hat{f}_m(x)x\, dx \leq C\frac{\sqrt{m}}{n},$$

where the constant C *only depends on the kernel* K *and not on the unknown density* f*.*

($C = \sum_{i=1}^L |\alpha_i| \sum_{i=1}^L |\alpha_i|/\sqrt{2\pi v_i}(1 + o_m(1))$). As the problem here only deals with densities on $\mathbb{R}^+$, the usual Nikol'skii class as defined in Tsybakov (2009) (see also Goldenschluger and Lepski (2010)) is not adequate. We need define a new class fitted with the bias term.

Definition 4.1. *For* $\beta \geq 0$ *and* ϕ *a nonnegative convex function on* $(0, +\infty)$ *such that* $\phi(1) \neq 0$*, consider the class of functions* f *defined on* $\mathbb{R}^+$ *having a derivative on* $(0, +\infty)$ $f^{(\ell)}$ *with* $\ell = \lfloor \beta \rfloor$ *satisfying, for all positive* u*,*

$$\left[\int_0^{+\infty} (u\, x)^{2\ell} \left(f^{(\ell)}(ux) - f^{(\ell)}(x)\right)^2 u\, dx\right]^{1/2} \leq \phi(u)|u-1|^{\beta-\ell}.$$

We denote this class by $\mathcal{H}_+(\beta, \phi)$*.*

Proposition 4.2. *(Bias bound) Assume that* $K_m = \sum_{i=1}^L \alpha_i K_m^{(i)}$ *where* $\sum_{i=1}^L \alpha_i = 1$*, for* $i = 1, \dots, L$*,* $K^{(i)}$ *satisfies Assumptions* (H1)-(H3) *(with variance denoted by* v_i*) and* $\int |u-1|^\beta K^{(j)}(u)du < +\infty$*. Assume that the density* f *belongs to the class* $\mathcal{H}_+(\beta, \phi)$*. Assume that the kernel* K *is of order* $\ell = \lfloor \beta \rfloor$ *(see (3.1)) and satisfies: there exists* m_0 *such that*

$$\sup_{m \geq m_0} m^{\beta/2} \int \psi(u)|u-1|^\beta |K_m(u)|du = \tilde{C} < +\infty, \tag{20}$$

with $\psi(u) = (\phi(1) + \phi(u))(1 + \frac{1}{u^{\ell+(1/2)}})$*. Then, for* $m \geq m_0$*,*

$$\int_0^{+\infty} \left(\mathbb{E}\hat{f}_m(x) - f(x)\right)^2 dx \leq \frac{\tilde{C}^2}{\ell!^2 m^\beta}.$$

Let us make some comments on the class $\mathcal{H}_+(\beta, \phi)$ and the assumption (20). The new class requires an enveloppe function ϕ (see the examples below). As $\phi(1) = \psi(1) \neq 0$, and ϕ is continuous by the convexity, we know that, for all i, as m tends to infinity, with $\bar{U}_{m,i}$ having density $K_m^{(i)}$,

$$m^{\beta/2}|\bar{U}_{m,i} - 1|^\beta \psi(\bar{U}_{m,i}) \to_{\mathcal{D}} v_i^{\beta/2}|Z|^\beta \phi(1),$$

where Z is a standard Gaussian variable. Condition (20) requires that this convergence in distribution implies the convergence of the expectation

$$\mathbb{E}m^{\beta/2}|\bar{U}_{m,i} - 1|^\beta \psi(\bar{U}_{m,i})$$

to the corresponding expectation $v_i^{\beta/2}\phi(1)\mathbb{E}|Z|^\beta$. Hence, the condition is not stringent.
Finally, we can state:

Theorem 4.1. *Under the assumptions of Propositions 4.1 and 4.2, the weighted MISE satifies*

$$\mathbb{E}\int_0^{+\infty} \left(\hat{f}_m(x) - f(x)\right)^2 w(x)dx \leq C\frac{\sqrt{m}}{n} + \frac{\tilde{C}^2}{\ell!^2 m^\beta}.$$

4.2. **Examples.**

(1) Consider the density $f(x) = e^{-x}1_{\mathbb{R}^+}(x)$. As a function on $\mathbb{R}$, it is not derivable. We have

$$\int (f(x+t) - f(x))^2 dx = (1 - e^{-|t|}) \leq |t|.$$

Hence, it belongs to the usual Nikol'skii class $\mathcal{H}(\beta, L)$ with $\beta = 1/2, L = 1$ (see *e.g.* Tsybakov, 2009). Thus, if we estimate f using a standard kernel estimator, the better rate of convergence of the risk is $n^{-2/3}$ (for the optimal bandwidth $h = n^{-1/3}$). As a function on $(0, +\infty)$, f is infinitely derivable with $f^{(\ell)}(x) = (-1)^\ell e^{-x}$. We have, for all integer ℓ

$$\int_0^{+\infty} (u\,x)^{2\ell} \left(f^{(\ell)}(ux) - f^{(\ell)}(x)\right)^2 u\,dx = \frac{\Gamma(2\ell+1)}{2^{2\ell+1}(1+u)^{2\ell+1}} P(u),$$

where the polynomial

$$P(u) = (1 + u^{2\ell+1})(1+u)^{2\ell+1} - 2(2u)^{2\ell+1}$$

satisfies $P(1) = P'(1) = 0, P''(1) \neq 0$. Hence, $P(u) = (u-1)^2 Q(u)$ where $Q(u)$ is a polynomial of degree 4ℓ such that $Q(1) \neq 0$. Thus f belongs to $\mathcal{H}_+(\beta, Q^{1/2})$ for all β. The optimal rate of convergence of the risk for the estimation of f can be $n^{-2\beta/(2\beta+1)}$ for all positive β (for $m = n^{-2/(2\beta+1)}$).

(2) For $f(x) = xe^{-x}1_{\mathbb{R}^+}(x)$, we have $\ell = 0$ and $\beta = 1$ for the usual Nikol'skii class as:

$$\int (f(x+t) - f(x))^2 dx = \frac{1}{2}(1 - e^{-|t|} - |t|e^{-|t|}) \leq t^2.$$

On $(0, +\infty)$, f is infinitely derivable with $f^{(\ell)}(x) = (-1)^\ell (x - \ell)e^{-x}1_{(0,+\infty)}(x)$. For all integer ℓ

$$\int_0^{+\infty} (u\,x)^{2\ell} \left(f^{(\ell)}(ux) - f^{(\ell)}(x)\right)^2 u\,dx = \frac{(\ell+1)\Gamma(2\ell+1)}{(1+u)^{2\ell+3}} R(u),$$

where

$$R(u) = \frac{1}{2^{2\ell+2}}(1 + u^{2\ell+1})(1+u)^{2\ell+3} - 2u^{2\ell+1}\left((2\ell+2)u + \ell(1+u^2)\right)$$

satisfies $R(1) = R'(1) = 0, R''(1) \neq 0$. Here again, $R(u) = (u-1)^2 L(u)$ where $L(u)$ is a polynomial of degree $4\ell + 2$ such that $L(1) \neq 0$. Thus f belongs to $\mathcal{H}_+(\beta, L^{1/2})$ for all β.

(3) For $f(x) = 1_{[0,1]}(x)$, we have $\ell = 0$ and

$$\int (f(x+t) - f(x))^2 dx = 2|t| \wedge 1,$$

$$\int_0^{+\infty} (f(ux) - f(x))^2\, u\,dx = |u - 1|.$$

Hence, $\beta = 1/2$ for both criteria.

So far, in the above examples, computations rely on the explicit expression of the density f. We can give a general property.

Proposition 4.3. *Assume that the density f is C^∞ on $(0, +\infty)$ and compactly supported. Then, for all integer ℓ, such that $f^{(\ell+1)} \neq 0$*

$$F(u) = \left[\int_0^{+\infty} (u\,x)^{2\ell} \left(f^{(\ell)}(ux) - f^{(\ell)}(x)\right)^2 u\,dx\right]^{1/2} \leq c(1 + u^{\ell+(1/2)})|u - 1|.$$

4.3. **Global bandwidth selection.** Let $\mathcal{M}_n$ be defined by (13) and consider again $K_m = \sum_{j=1}^{L} K_m^{(j)}$ an extended kernel as defined in (3)-(4) where the densities $K^{(j)}$ satisfy (H1)-(H3). We still consider $\hat{f}_{m,m'}$ defined by (14).

Now, define

$$V(m) = \kappa C(K)\frac{\sqrt{m}}{n}$$

where κ is a numerical constant and $C(K)$ defined by (17),

$$A(m) = \sup_{m' \in \mathcal{M}_n} \left(\|\hat{f}_{m'} - \hat{f}_{m,m'}\|_w^2 - V(m') \right)_+ . \tag{21}$$

The adaptive estimator is then $\tilde{f} = \hat{f}_{\hat{m}}$ with

$$\hat{m} = \arg\min_{m \in \mathcal{M}_n} \left(A(m) + V(m) \right).$$

We can prove the following result

Theorem 4.2. *Assume that f belongs to $\mathbb{L}^2((0,+\infty))$. Under the assumptions of Proposition 4.1, we have*

$$\mathbb{E}(\|\tilde{f} - f\|_w^2) \leq C \inf_{m \in \mathcal{M}_n} \Big(\int_0^{+\infty} (\mathbb{E}(\hat{f}_m(x) - f(x))^2 dx + V(m)) + \frac{C'}{n}.$$

Corollary 4.1. *Assume that f belongs to $\mathbb{L}^2((0,+\infty))$. Under the assumptions of Propositions 4.1 and 4.2, with $\mathcal{M}_n$ defined in (13),*

$$\mathbb{E}(\|\tilde{f} - f\|_w^2) \leq O(n^{-2\beta/(2\beta+1)}).$$

The corollary follows immediately from Theorem 4.2 and Proposition 4.2. It is worth noticing that the optimal rate of convergence is automatically reached by the adaptive estimator.

5. Practical implementation.

5.1. **The benchmark: standard kernel estimators with local adaptive bandwidth selection.** We compare our method with the one, based on iterated kernels, described in Goldenschluger and Lepski (2010), transposed in a pointwise version. Note that the idea to use several kernels was introduced by Devroye (1989). To be more precise, let

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x), \quad K_h(u) = \frac{1}{h}K(\frac{u}{h}),$$

where K is a symmetric kernel on $\mathbb{R}$, $\int_{\mathbb{R}} |K(u)|du < +\infty$,$\int_{\mathbb{R}} K(u)du = 1$ and $\int_{\mathbb{R}} K^2(u)du < +\infty$. Assume that f is bounded and belongs to a Hölder class $\Sigma(\beta, C)$. Let K be a square integrable kernel of order β in the usual sense (*i.e.* $\int_{\mathbb{R}} u^k K(u)du = 0$ for $k = 1, \ldots, \ell$, where $\ell = \lfloor \beta \rfloor$ and $\int_{\mathbb{R}} |u|^\beta |K(u)|du < +\infty$). Then a classical squared bias-variance decomposition of the pointwise risk gives a bound $C_\beta h^{2\beta} + (\|f\|_\infty \int K^2)/(nh)$ and a rate of order $n^{-2\beta/(2\beta+1)}$ if h is chosen proportional to $n^{-1/(2\beta+1)}$, see Tsybakov (2009).

In Goldenschluger and Lespki (2010), only the global bandwidth selection is studied. Let us describe how local automatic bandwidth selection can be performed. We define

$$\hat{f}_{h',h}(x) = K_{h'} \star \hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h'} \star K_h(X_i - x) = f_{h,h'}(x)$$

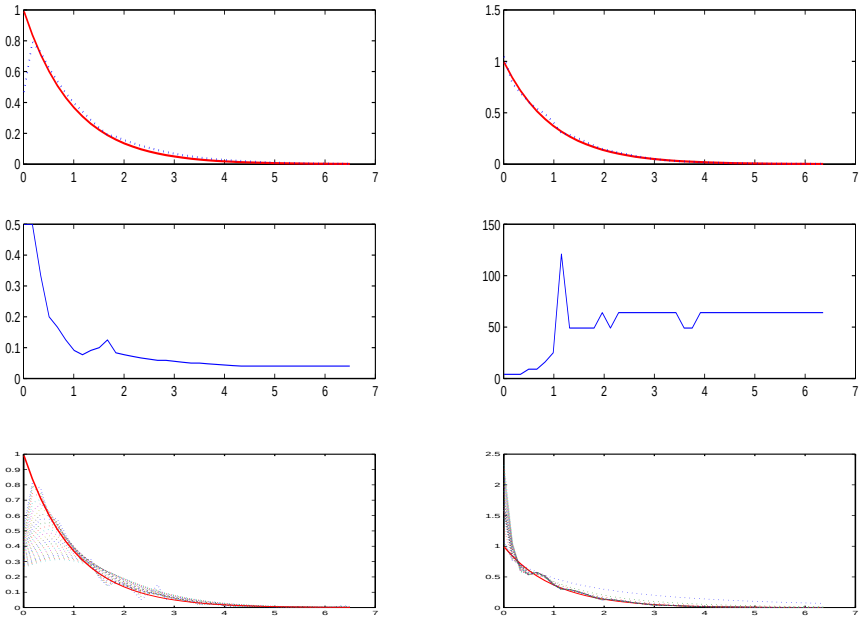


FIGURE 1. Exponential $\mathcal{E}(1)$ density. True function (bold) and estimation for one path (dotted) with $n = 1000$, Lepski method (up left), our method (up right), selected h (middle left) or m (middle right), true function (bold) and set of all the estimators proposed to the selection algorithm (bottom).

and for $\mathcal{H}_n = \{h_k, k = 1, \dots, M_n\}$, $M_n \leq n$, we set

$$S(h) = \kappa'(\int_{\mathbb{R}} |K|)^2(\int_{\mathbb{R}} K^2(u)du)\frac{\log(n)}{nh}, \quad B(x,h) = \sup_{h' \in \mathcal{H}_n} \left((\hat{f}_{h,h'}(x) - \hat{f}_{h'}(x))^2 - S(h')\right)_+$$

and

$$\hat{h}(x) = \arg\min_{h \in \mathcal{H}_n} \left(B(x,h) + S(h)\right).$$

If f is bounded and in a Hölder class $\Sigma(\beta, C)$ and K is a square integrable kernel of order $\ell \geq \beta$, we can prove that, for n large enough, $\mathbb{E}[(\hat{f}_{\hat{h}(x)}(x) - f(x))^2] = O((\log(n)/n)^{2\beta/(2\beta+1)})$.

As the steps are analogous to the ones used to prove Theorem 3.1, we omit the proof. In practice, Gaussian kernels are used to ease the computation of $K_h \star K_{h'}$.

5.2. **Results.** We provide in this section results of simulation experiments. Concretely, the penalty is computed using the true value of $\|f\|_\infty$, $\kappa = 5.10^{-3}$ and $\kappa' = 0.1$. We choose $\mathcal{H}_n = \{h_k = k/M, k = 1, \dots, M\}$ and $\mathcal{M}_n = \{m_k = k^2, k = 1, \dots, M\}$ with $M = 25$. The function is estimated on $[a, b]$ with $a = \max(\min(X), 0.01\max(X))$ and $b = \max(X)$ where $X = (X_1, \dots, X_n)$ is the vector of observations. Estimators are plotted for 40 equispaced points of $[a, b]$. We take one Inverse Gaussian kernel for K_m with $a = 1$ ($L = 1$) and one standard Gaussian kernel for K_h. Then, we know in all cases the convoluted functions and the terms $\int |K|$ or $\int K^2$.

We show in Figures 1-2 the variability of the choices of h or m depending on the point of estimation. The upper plots give the true function and one estimation and the middle ones give the selected m or h at the corresponding points. It is clear that both methods adapt very well to the form of the function. The bottom plots give the curves for all values of h and m. Figure

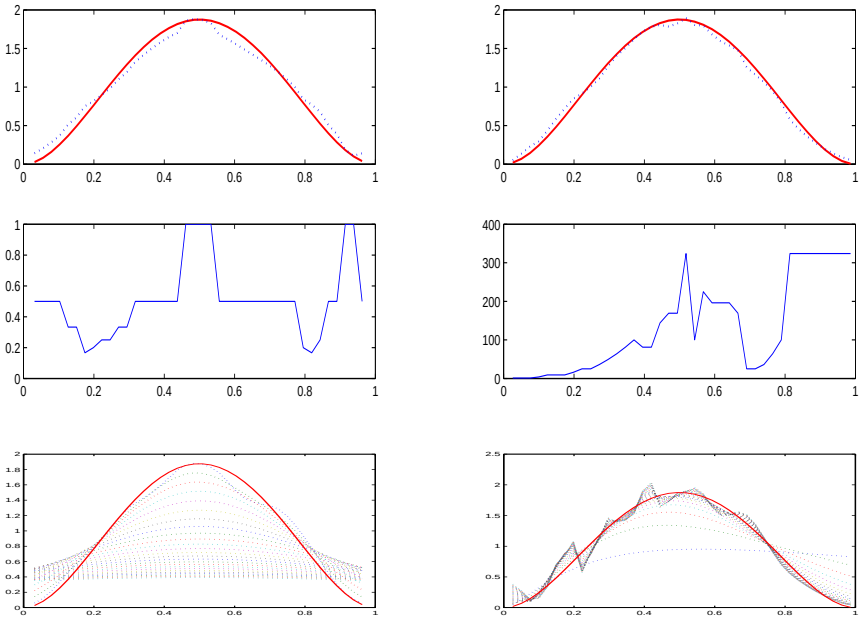


FIGURE 2. Beta $\beta(3,3)$ density. True function (bold) and estimation for one path (dotted) with $n = 1000$, Lepski method (up left), our method (up right), selected h (middle left) or m (middle right), true function (bold) and set of all the estimators proposed to the selection algorithm (bottom).

1 shows that, near 0, the two methods react differently. For all proposals, the usual kernel estimators are too low. On the contrary, for our kernels which are fitted to nonnegative data, the proposed estimators take larger values.

We study then the compared variability of the methods in Figure 3 for four different functions: an exponential $\mathcal{E}(1)$, a beta $\beta(1,3)$, a beta $\beta(3,3)$ and a lognormal with $\mu = -1$ and $\sigma = 0.25$. We also considered two different samples sizes $n = 400$ and $n = 1000$ for the exponential and the lognormal. We observe the same difference of behaviours near 0 as previously noticed for the exponential $\mathcal{E}(1)$ and the beta $\beta(1,3)$. This confirms the fact that our method performs better from this point of view.

6. Discussion and concluding remarks.

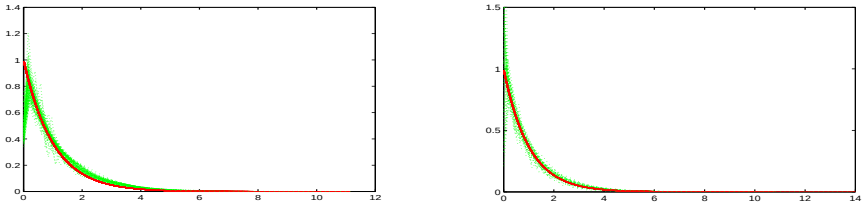
In this paper, we propose a new type of kernel estimators fitted to nonnegative data, based on kernels which are approximations of δ_1. In usual kernel methods, the intuition is that the estimation at x counts the number of observations X_k such that $X_k - x$ is close to 0. In our strategy, the intuition is that the estimator at x counts the number of observations X_k such that X_k/x is close to 1.

We prove that we can develop the complete theory of nonparametric density estimation for nonnegative data, from both the pointwise and the global point of view: bias-variance decomposition, higher order kernels for fitting functional regularities, adaptive bandwidth selection.

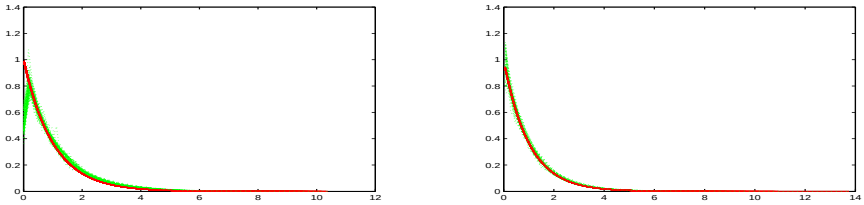
As noted above, $\int_0^{+\infty} \hat{f}_m(x)dx = 1 + O(1/m)$. If we consider:

$$\check{f}_m(x) = \frac{1}{nx} \sum_{k=1}^{n} \frac{X_k}{x} K_m\left(\frac{X_k}{x}\right)$$

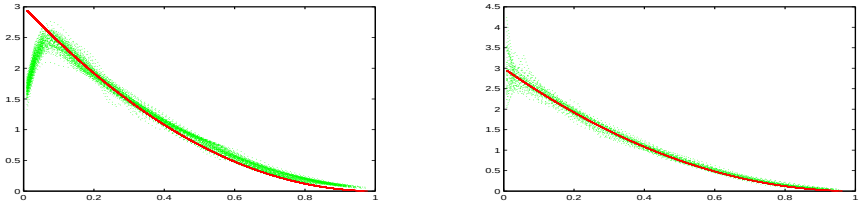
Exponential $\mathcal{E}(1)$ distribution, $n = 400$



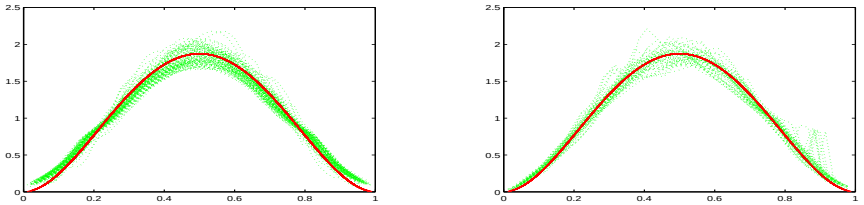
Exponential $\mathcal{E}(1)$ distribution, $n = 1000$



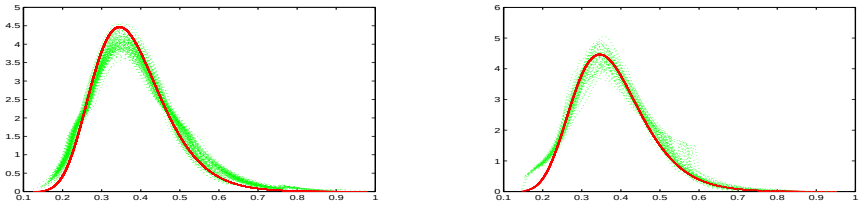
Beta $\beta(1, 3)$ distribution, $n = 1000$



Beta $\beta(3, 3)$ distribution, $n = 1000$



Lognormal $\mu = -1, \sigma = 0.25$ distribution, $n = 400$



Lognormal $\mu = -1, \sigma = 0.25$ distribution, $n = 1000$

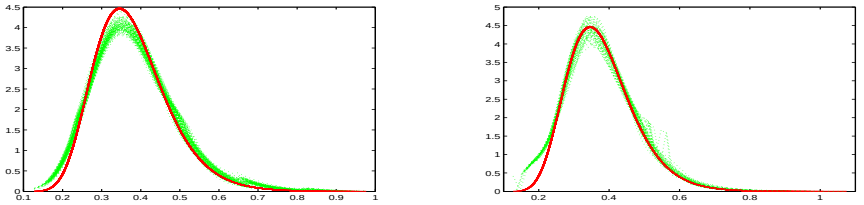


FIGURE 3. True density (bold) and 50 estimated densites (dotted), Lepski method (left), our method (right).

then $\int_0^{+\infty} \check{f}_m(x)dx = 1$. The theoretical study and the numerical performances are analogous to those of $\hat{f}_m$. But, proofs for the adaptive procedure are much more difficult, especially for what concerns the Bernstein and the Talagrand inequalities (unbounded empirical processes).

In our numerical simulations, we have implemented only the pointwise adaptive method as the results are really convincing. To compare our estimators to more standard kernel estimators, we have described how the adaptive method of Goldenschluger and Lespki (2010) can be done locally. The numerical results show that both methods perform well, with a slight advantage for our method near 0 when $f(0) > 0$.

It is worth stressing that we have at disposal explicit formulae for the iterated kernel $K_m \odot K_{m'}$ with $K = IG(1,1)$, involving a Bessel function available in Matlab. Note that some numerical problems may appear in this computation due to the very small values of the Bessel function $\tilde{K}_0(z)$ for large z. An interesting alternative is to use the formula

$$e^z \tilde{K}_0(z) = \int_0^{+\infty} e^{-zs/2} \frac{ds}{\sqrt{s(s+4)}},$$

which is obtained from (8) and numerically more stable.

7. Proofs

7.1. **Proof of Lemma 2.1.** Recall that the Fourier transform of K_m is $K_m^*(t) = (K^*(t/m))^m$. Thus, by the change of variables $t = s\sqrt{m}$, we get:

$$\int_{\mathbb{R}} |K_m^*(t)|^\alpha dt = \sqrt{m} \int_{\mathbb{R}} |K^*(s/\sqrt{m})|^{\alpha m} ds.$$

We now prove that, for $\alpha \geq 1$, as m tends to infinity:

$$D_m = \int_{\mathbb{R}} |K^*(s/\sqrt{m})|^{\alpha m} ds - \int_{\mathbb{R}} e^{-\alpha v s^2/2} ds \to 0 \tag{22}$$

where $\int_{\mathbb{R}} e^{-\alpha v s^2/2} ds = \sqrt{2\pi/\alpha v}$. First, note that $\varphi(t) = K^*(t)e^{-it}$ is the characteristic function of a distribution which is centered, has the same variance v as K and satisfies $|\varphi(t)| = |K^*(t)|$ for all t. Hence, we can replace K^* by φ in (22) to study D_m. Now, we can follow closely the proof given in Feller (1971, p.516). We have

$$\varphi(t) = 1 - \frac{vt^2}{2} + o(t^2).$$

Using that

$$\varphi(s/\sqrt{m}) = 1 - \frac{vs^2}{2m} + o(1/m),$$

we standardly deduce that, uniformly on compact sets,

$$v_m(s) := |\varphi(s/\sqrt{m})|^{\alpha m} - e^{-\alpha v s^2/2} \to 0.$$

Moreover, we can choose $\delta > 0$ such that, for $|t| \leq \delta$,

$$|\varphi(t)| \leq e^{-t^2 v/4}.$$

Therefore, we split the integral into three terms:

$$\int_{\mathbb{R}} |v_m(s)| ds = \int_{|s| \leq a} \ldots + \int_{a < |s| < \delta\sqrt{m}} \ldots + \int_{|s| \geq \delta\sqrt{m}} \ldots \tag{23}$$

For the second term, we use:

$$\int_{a<|s|<\delta\sqrt{m}} |v_m(s)|ds \leq 2\int_{a<|s|<\delta\sqrt{m}} e^{-s^2\alpha v/2}ds.$$

For all $\varepsilon > 0$, this term can be made smaller than ε by choosing a large enough. The first term in (23) tends to 0. For the third term, we know that, for $t \neq 0$, $|\varphi(t)| < 1$ and $\varphi(t)$ tends to 0 as t tends to infinity. Therefore, $\eta = \sup_{|t|\geq\delta} |\varphi(t)| < 1$. Hence, we can write for $m \geq 2/\alpha$,

$$\int_{|s|\geq\delta\sqrt{m}} |v_m(s)|ds \leq \eta^{(\alpha m-2)} \int |\varphi(s/\sqrt{m})|^2 ds + \int_{|s|\geq\delta\sqrt{m}} e^{-s^2\alpha v/2}ds.$$

As $\int |\varphi(s/\sqrt{m})|^2 ds = \sqrt{m}\int |\varphi(t)|^2 dt$, and $\int |\varphi(t)|^2 dt < \infty$ by the assumption that K is in L^2, we get (22) and the result follows.
Using Fourier inversion and the Parseval equality yield:

$$K_m(u) \leq \frac{1}{2\pi}\int_{\mathbb{R}} |K_m^*(t)|dt \quad \text{and } ||K_m||_2^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |K_m^*(t)|^2 dt,$$

Thus, the first point.
For the second point, we use analogously that

$$< K_m^{(1)}, K_m^{(2)} >= \frac{1}{2\pi} < (K_m^{(1)})^*, (K_m^{(2)})^* > .$$

Then, we prove in the same way

$$\frac{1}{\sqrt{m}} < (K_m^{(1)})^*, (K_m^{(2)})^* > - \int_{\mathbb{R}} e^{-(v_1+v_2)s^2/2}ds \to 0.$$

□

7.2. **Proof of Proposition 3.1.** We have:

$$\mathrm{Var}\hat{f}_m(x) = \frac{1}{nx^2}\left[\int_0^{+\infty} K_m^2(t/x)f(t)dt - \left(\int_0^{+\infty} K_m(t/x)f(t)dt\right)^2\right] \leq \frac{||f||_\infty}{nx}\int_0^{+\infty} K_m^2(u)du.$$

We develop K_m^2. By Lemma 2.1, as m tends to infinity, for $1 \leq i, j \leq L$,

$$\int_0^{+\infty} K_m^{(i)}(u)K_m^{(j)}(u)du = \sqrt{m}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v_i+v_j}}(1+o(1)).$$

The result follows. □

7.3. **Proof of Lemma 7.1.** For the proof, we omit the superscript (j) and assume that $K^{(j)} = K$ is a density satisfying the assumptions. Consider first the case $0 < \alpha < 2$. Then, by the Holder inequality, as $4/\alpha > 1$,

$$\nu_{\alpha,m} = \mathbb{E}|\bar{U}_m - 1|^\alpha \leq \left(\mathbb{E}(\bar{U}_m - 1)^4\right)^{\alpha/4}.$$

Some elementary computations lead to

$$\mathbb{E}(\bar{U}_m - 1)^4 = \frac{\mu_4}{m^3} + 6v^2\frac{m-1}{m^3}.$$

As $\alpha/4 < 1/2 < 1$,

$$\left(\mathbb{E}(\bar{U}_m - 1)^4\right)^{\alpha/4} \leq (6v^2)^{\alpha/4}\frac{1}{m^{\alpha/2}} + \frac{\mu_4^{\alpha/4}}{m^{3\alpha/4}}$$

For $\alpha \geq 2$, we can use the Rosenthal inequality (see *e.g.* Hall and Heyde, 1980, p.23). For some constant $c(\alpha)$,

$$\nu_{\alpha,m} \leq c(\alpha)\left(\frac{v^\alpha}{m^{\alpha/2}} + \frac{\nu_\alpha}{m^{\alpha-1}}\right),$$

where $\alpha - 1 \geq \alpha/2$. □

7.4. **Proof of Proposition 3.2.** Before proving the result, we evaluate the rate of absolute centered moments of a density $K_m^{(j)}$ when $K^{(j)}$ satisfies (H1)-(H3). Let us set, for ℓ integer, $\gamma > 0, m \geq 2$,

$$\mu_\ell^{(j)} = \int_0^{+\infty} (u-1)^\ell K^{(j)}(u)du, \quad \mu_{\ell,m}^{(j)} = \int_0^{+\infty} (u-1)^\ell K_m^{(j)}(u)du, \tag{24}$$

$(\mu_1^{(j)} = \mu_{1,m}^{(j)} = 0, \mu_2^{(j)} = v_j, \mu_{2,m}^{(j)} = v_j/m)$,

$$\nu_\gamma^{(j)} = \int_0^{+\infty} |u-1|^\gamma K^{(j)}(u)du, \quad \nu_{\gamma,m}^{(j)} = \int_0^{+\infty} |u-1|^\gamma K_m^{(j)}(u)du. \tag{25}$$

Lemma 7.1. *Assume that the density $K^{(j)}$ satisfies* (H1) *and* (H3). *Then, for all $\alpha \leq \gamma$,*

$$\nu_{\alpha,m}^{(j)} \leq \frac{c_j(\alpha)}{m^{\alpha/2}},$$

where the constant $c_j(\alpha)$ depends on $\alpha, v_j, \mu_4^{(j)}$ and $\nu_\alpha^{(j)}$.

The moments $\mu_{\ell,m}^{(j)}$ for ℓ integer can be computed and expressed as explicit functions of the $\mu_k^{(j)}, k \leq \ell$. In particular, if $\mu_{2\ell}^{(j)} < +\infty$ for some integer ℓ, then, $\mu_{2\ell,m}^{(j)} \leq m^{-\ell} C(\mu_{2k}^{(j)}, k \leq \ell)$.

Now we prove Prosposition 3.2. By the Taylor formula, and the assumption on K_m,

$$\mathbb{E}\hat{f}_m(x) - f(x) = x^\ell \int_0^{+\infty} du(u-1)^\ell K_m(u) \int_0^1 \frac{s^{\ell-1}}{(\ell-1)!}\left(f^{(\ell)}(x + sx(u-1))) - f^{(\ell)}(x)\right) ds.$$

The result follows using the assumption on f and Lemma 7.1 (see Tsybakov (2009)).□

7.5. **Proof of Proposition 3.4.** Define

$$\varphi(t) = K^*(t)e^{-it} = e^{\psi(t)},$$

which is the characteristic function of $V = U - 1$ for U a random variable with density K. The following relations hold:

$$i\kappa_1 = \varphi'(0) = \psi'(0) = 0, \quad -\kappa_2 = \varphi''(0) = \psi''(0) = -v,$$

and for $j \geq 3$,

$$\varphi^{(j)}(0) = i^j \mu_j = i^j \mathbb{E}(V^j), \quad \psi^{(j)}(0) = i^j \int_0^{+\infty} x^j n(x)dx.$$

Thus, the first cumulant of V is $\kappa_1 = 0$ and for $j \geq 2$, the j-th cumulant of V is

$$\kappa_j = \int_0^{+\infty} x^j n(x)dx,$$

($\kappa_2 = v$). As K^* is a characteristic function, and K belongs to $L^2(\mathbb{R}^+)$, we have, for $a > 2$:

$$|K_a^*(t)| \leq |K^*(\frac{t}{a})|^2 \quad \text{and } |K_a^*(t)|^2 \leq |K^*(\frac{t}{a})|^2$$

which implies that the distribution P_a has a density K_a belonging to $L^2(\mathbb{R}^+)$.
Let us now establish the links between the moments $\mu_{j,a} = \mathbb{E}(V_a^j)$ of $V_a = U_a - 1$ where U_a has density K_a and the moments μ_j of $V = U - 1$ where U has density $K = K_1$. We denote by $\kappa_{j,a}$ the j-th cumulant of V_a and set

$$\varphi_a(t) = e^{\psi_a(t)}.$$

We have

$$\psi_a'(0) = 0 = \varphi_a'(0). \tag{26}$$

For $j \geq 2$,

$$\psi_a^{(j)}(0) = i^j \kappa_{j,a} = i^j \int_0^{+\infty} x^j n_a(x) dx = i^j \frac{1}{a^{j-1}} \kappa_j. \tag{27}$$

Using (26) and tedious computations, we obtain:

$$\varphi_a''(0) = \psi_a''(0), \quad \varphi_a^{(3)}(0) = \psi_a^{(3)}(0),$$
$$\varphi_a^{(4)}(0) = \psi_a^{(4)}(0) + 3(\psi_a''(0))^2, \quad \varphi_a^{(5)}(0) = \psi_a^{(5)}(0) + 9\psi_a''(0)\psi_a^{(3)}(0).$$

This implies:

$$\mu_{2,a} = v_a = v/a, \mu_{3,a} = \kappa_3/a^2, \mu_{4,a} = \kappa_4/a^3 + 3v^2/a^2, \mu_{5,a} = \kappa_5/a^4 + 9v\kappa_3/a^3.$$

More generally, cumulants and (centered) moments are linked by the induction formula (see *e.g.* Abramowitz and Stegun(1964)) (using that $\kappa_1 = 0$, see (27)):

$$\kappa_{n,a} = \mu_{n,a} - \sum_{k=2}^{n-2} \binom{n-1}{k-1} \kappa_{k,a} \mu_{n-k,a}.$$

We deduce, by a tedious but elementary induction:

$$\mu_{n,a} = \sum_{\ell=k(n)}^{n-1} \frac{c(\ell)}{a^\ell}$$

where $c(n-1) = \kappa_n$, the other constants $c(\ell)$ depend on the cumulants $\kappa_\ell, 2 \leq \ell \leq n-1$ of K and $k(n) = n/2$ for n even, $k(n) = (n+1)/2$ for n odd. □

7.6. **Proof of Proposition 3.5.** By Proposition (3.4), for $a > 2\ell$, the distributions $P_{aj^{-1}}$ with characteristic functions specified by the relation (11) have a density satisfying (H1) and admit moments up to order ℓ. Hence the functions $K^{(j)}$ are well-defined.
First, we have:

$$\sum_{j=1}^{\ell} \binom{\ell}{j} (-1)^{j+1} = (-1)[(1-1)^\ell - 1] = 1.$$

Then, we have to check that, for $k = 1, \ldots, \ell$, $\int_0^{+\infty} (u-1)^k K_m(u) du = 0$. By Proposition (3.4),

$$\int_0^{+\infty} (u-1)^k K_{amj^{-1}}(u) du = \sum_{i \leq k-1} \frac{j^i}{(am)^i} c(i),$$

where $c(i) = 0$ for $i < k/2$ for even k, and $c(i) = 0$, for $i < (k+1)/2$ when k is odd. Hence, it is enough to prove that, for $k = 1, \ldots, \ell - 1$

$$\sum_{j=1}^{\ell} \binom{\ell}{j} (-1)^{j+1} j^k = 0,$$

which, in turn, holds as, for $k = 1, \dots, \ell - 1$,

$$\begin{aligned}\sum_{j=1}^{\ell} \binom{\ell}{j} (-1)^{j+1} j(j-1)\dots(j-k+1) &= \sum_{j=k}^{\ell} \binom{\ell}{j} (-1)^{j+1} j(j-1)\dots(j-k+1) \\ &= \frac{\ell!}{(\ell-k)!} \sum_{j=k}^{\ell} \binom{\ell-k}{j-k} (-1)^{j+1} \\ &= (-1)^{k+1}(1-1)^{\ell-k} = 0.\end{aligned}$$

□

7.7. **Proof of Theorem 3.1.**

7.7.1. *Three useful Lemmas.* Let us set, for $m, m' > 0$,

$$B_m f(x) = \mathbb{E}\hat{f}_m(x) - f(x), \quad B_{m,m'} f(x) = \mathbb{E}\hat{f}_{m,m'}(x) - f(x). \tag{28}$$

The following relation between bias terms holds.

Lemma 7.2.

$$B_{m,m'} f(x) = B_{m'} f(x) + \int_0^{+\infty} K_{m'}(u) B_m f(xu) du.$$

Proof. Let

$$\hat{f}_{m,i}(x) = \frac{1}{nx} \sum_{k=1}^{n} K_m^{(i)}(\frac{X_k}{x}), \quad \hat{f}_{m,m',i,j}(x) = \frac{1}{nx} \sum_{k=1}^{n} K_m^{(i)} \odot K_{m'}^{(j)}(\frac{X_k}{x}),$$

$$B_{m,i} f(x) = \mathbb{E}\hat{f}_{m,i}(x) - f(x), \quad B_{m,m',i,j} f(x) = \mathbb{E}\hat{f}_{m,m',i,j}(x) - f(x).$$

We have:

$$\begin{aligned}\mathbb{E}\hat{f}_{m,m',i,j}(x) &= \int\int K_m^{(i)}(t/xy) K_{m'}^{(j)}(y) f(t) \frac{dydt}{xy} \\ &= \int K_{m'}^{(j)}(y) \frac{dy}{xy} \int xy K_m^{(i)}(v) f(xyv) dv \\ &= \int K_{m'}^{(j)}(y) [\mathbb{E}\hat{f}_{m,i}(xy) - f(xy)] dy + \int K_{m'}^{(j)}(y) f(xy) dy.\end{aligned}$$

Using $\sum_{i=1}^{L} \alpha_i = 1$ yields,

$$\sum_{i=1}^{L} \alpha_i \mathbb{E}\hat{f}_{m,m',i,j}(x) = \int K_{m'}^{(j)}(y) B_m f(xy) dy + B_{m',j} f(x) + f(x).$$

And

$$B_{m,m'} f(x) = \int (\sum_{j=1}^{L} \alpha_j K_{m'}^{(j)}(y)) B_m f(xy) dy + \sum_{j=1}^{L} \alpha_j B_{m',j} f(x),$$

which is the result.□

Lemma 7.3. *Let $K_m = \sum_{i=1}^{L} \alpha_i K_m^{(i)}$ with $\sum_{i=1}^{L} \alpha_i = 1$ be an extended kernel where the densities $K^{(j)}$ satisfy* (H1)-(H3)*. Then,*

$$||K_m \odot K_{m'}||_\infty \leq \sqrt{m \wedge m'} (\sum_{i=1}^{L} |\alpha_i|)(\sum_{i=1}^{L} \frac{|\alpha_i|}{\sqrt{2\pi v_i}})(1 + o(1)). \tag{29}$$

Proof. Consider $U_k^{(i)}, k = 1, \ldots, m, V_\ell^{(j)}, \ell = 1, \ldots, m'$ independent random variables where the $U_k^{(i)}$'s have density $K^{(i)}$ and the $V_\ell^{(j)}$'s have density $K^{(j)}$. Recall that $K_m^{(i)}$ is the density of $\bar{U}_m^{(i)} = (U_1^{(i)} + \ldots + U_m^{(i)})/m$ and $I_m^{(i)} = \mathbb{E}(1/\bar{U}_m^{(i)}) = 1 + O(1/m)$. Analogously, $K_{m'}^{(j)}$ is the density of $\bar{V}_{m'}^{(j)} = (V_1^{(j)} + \ldots + V_{m'}^{(j)})/m'$ and $C_{m'}^{(j)} = \mathbb{E}(1/\bar{V}_{m'}^{(j)}) = 1 + O(1/m)$. We have

$$K_m \odot K_{m'}(u) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iut}(K_m \odot K_{m'})^*(t)dt = \frac{1}{2\pi}\sum_{1\le j,\ell\le L}\alpha_j\alpha_\ell \int_{\mathbb{R}} e^{-iut}(K_m^{(j)} \odot K_{m'}^{(\ell)})^*(t)dt,$$

and

$$(K_m^{(j)} \odot K_{m'}^{(\ell)})^*(t) = \mathbb{E}\left((K^{(\ell)})^*(\frac{t\bar{U}_m^{(j)}}{m'})\right)^{m'}.$$

Setting $t = s\sqrt{m'}/\bar{U}_m^{(j)}$, we get

$$\int_{\mathbb{R}} e^{-iut}(K_m^{(j)} \odot K_{m'}^{(\ell)})^*(t)dt = \sqrt{m'}\mathbb{E}\left(\frac{1}{\bar{U}_m^{(j)}}\int_{\mathbb{R}} e^{-ius\sqrt{m'}/\bar{U}_m^{(j)}}[(K^{(\ell)})^*(\frac{s}{\sqrt{m'}})]^{m'}ds\right).$$

Using that $\mathbb{E}(\frac{1}{\bar{U}_m^{(j)}}) = 1 + O(1/m)$, and Lemma 2.1, we get

$$||K_m^{(j)} \odot K_{m'}^{(\ell)}(u)||_\infty \le \sqrt{m'}(1 + O(1/m))(2\pi v_\ell)^{-(1/2)}(1 + o(1))$$

which yields the result by symmetry in m, m'. □

Lemma 7.4. *We have:*

$$Var\hat{f}_{m,m'}(x) \le \sqrt{m \wedge m'}\frac{C(K)||f||_\infty}{nx}(1 + o(1)).$$

where the constant $C(K)$ (see (17)) does not depend on the density f.

Proof. We have:

$$\mathrm{Var}\hat{f}_{m,m'}(x) \le \frac{1}{nx}\int_0^{+\infty} f(ux)\,(K_m \odot K_{m'}(u))^2\,du.$$

we conclude using Lemma 7.3 and the fact that $K_m^{(i)} \odot K_{m'}^{(j)}(u)$ are densities. □

7.7.2. *Proof of Theorem 3.1.* First note that the definition of $\hat{m}(x_0)$ given in (19) implies that $A(\hat{m}(x_0), x_0) + V(\hat{m}(x_0), x_0) \le A(m, x_0) + V(m, x_0)$ for all $m \in \mathcal{M}_n$. Hence, for m any element of $\mathcal{M}_n$, we can write the decomposition

$$\begin{aligned}(\tilde{f}(x_0) - f(x_0))^2 &\le 3[(\hat{f}_{\hat{m}(x_0)}(x_0) - \hat{f}_{m,\hat{m}(x_0)}(x_0))^2 + (\hat{f}_{m,\hat{m}(x_0)}(x_0) - \hat{f}_m(x_0))^2 + (\hat{f}_m(x_0) - f(x_0))^2]\\ &\le 3(A(m, x_0) + V(\hat{m}(x_0), x_0)) + 3(A(\hat{m}(x_0), x_0) + V(m, x_0)) + 3(\hat{f}_m(x_0) - f(x_0))^2\\ &\le 6(A(m, x_0) + V(m, x_0)) + 3(\hat{f}_m(x_0) - f(x_0))^2.\end{aligned}$$

Therefore,

$$\mathbb{E}[(\tilde{f}(x_0) - f(x_0))^2] \le 3\mathbb{E}[(\hat{f}_m(x_0) - f(x_0))^2] + 6V(m, x_0) + 6\mathbb{E}(A(m, x_0)).$$

Let us study $A(m, x_0)$ (see (18)). Let $\mathbb{E}(\hat{f}_m(x)) = f_m(x)$ and $\mathbb{E}(\hat{f}_{m,m'}(x)) = f_{m,m'}(x)$. Then $(\hat{f}_{m'}(x_0) - \hat{f}_{m,m'}(x_0))^2 \le 3(\hat{f}_{m'}(x_0) - f_{m'}(x_0))^2 + 3(\hat{f}_{m,m'}(x_0) - f_{m,m'}(x_0))^2 + 3(f_{m'}(x_0) - f_{m,m'}(x_0))^2$. By Lemma 7.2, for all $m, m' \in \mathcal{M}_n$,

$$|f_{m'}(x_0) - f_{m,m'}(x_0)| = |\int_0^{+\infty} B_m f(x_0 u)K_{m'}(u)du|.$$

Therefore, using that each $K_{m'}^{(i)}$ is a density, Proposition 7.1 and Proposition 3.2, we obtain:

$$\begin{aligned}
|f_{m'}(x_0) - f_{m,m'}(x_0)| &\leq \sum_{i=1}^{L} |\alpha_i| \int_0^{+\infty} |(B_m f)(xu)| K_{m'}^{(i)}(u) du \\
&\leq \frac{CC(\beta_0)}{\lfloor \beta_0 \rfloor !} \frac{x_0^{\beta_0}}{m^{\beta_0/2}} \sum_{i=1}^{L} |\alpha_i| \int u^{\beta} K_{m'}^{(i)}(u) du \\
&\leq \frac{CC(\beta_0) 2^{\beta_0}}{\lfloor \beta_0 \rfloor !} \frac{x_0^{\beta_0}}{m^{\beta_0/2}} \sum_{i=1}^{L} |\alpha_i| (\int K_{m'}^{(i)}(u) du + \nu_{\beta_0,m'}^{(i)}) \\
&\leq \frac{CC(\beta_0) 2^{\beta_0}}{\lfloor \beta_0 \rfloor !} \frac{x_0^{\beta_0}}{m^{\beta_0/2}} \sum_{i=1}^{L} |\alpha_i| (1 + c_i(\beta_0)(m')^{-\beta_0/2}) \leq C'(\beta_0, \alpha) \frac{x_0^{\beta_0} + 1}{m^{\beta_0/2}},
\end{aligned}$$

where

$$C'(\beta, \alpha) = \frac{CC(\beta_0) 2^{\beta_0}}{\lfloor \beta_0 \rfloor !} \sum_{i=1}^{L} |\alpha_i| (1 + c_i(\beta_0)).$$

Thus,

$$(f_{m'}(x_0) - f_{m,m'}(x_0))^2 \leq [C'(\beta_0, \alpha)]^2 \frac{2(x_0^{2\beta_0} + 1)}{m^{\beta_0}}.$$

We can write:

$$\begin{aligned}
A(m, x_0) &\leq 3 \sup_{m'} \left((\hat{f}_{m'}(x_0) - f_{m'}(x_0))^2 - \frac{V(m', x_0)}{6} \right)_+ \\
&\quad + 3 \sup_{m'} \left((\hat{f}_{m,m'}(x_0) - f_{m,m'}(x_0))^2 - \frac{V(m', x_0)}{6} \right)_+ + 6(C'(\beta, \alpha))^2 \frac{x_0^{2\beta} + 1}{m^{\beta}}.
\end{aligned}$$

Now, we can prove the following Lemmas:

Lemma 7.5. *Under the assumptions of Theorem 3.1, we have*

$$\mathbb{E}\left(\sup_{m'} \left((\hat{f}_{m'}(x_0) - f_{m'}(x_0))^2 - \frac{V(m', x_0)}{6} \right)_+ \right) \leq \frac{C}{n}.$$

Lemma 7.6. *Under the assumptions of Theorem 3.1, we have*

$$\mathbb{E}\left(\sup_{m'} \left((\hat{f}_{m,m'}(x_0) - f_{m,m'}(x_0))^2 - \frac{V(m', x_0)}{6} \right)_+ \right) \leq \frac{C''}{n}.$$

This yields that, $\forall m \in \mathcal{M}_n$,

$$\mathbb{E}((\tilde{f}(x_0) - f(x_0))^2) \leq 3\mathbb{E}((\hat{f}_m(x_0) - f(x_0))^2) + 6V(m, x_0) + 6(C'(\beta_0, \alpha))^2 \frac{x_0^{2\beta_0} + 1}{m^{\beta_0}} + \frac{6C''}{n}.$$

As $\mathbb{E}((\hat{f}_m(x_0) - f(x_0))^2) \lesssim V(m, x_0) + C_3 \frac{x_0^{2\beta_0}}{m^{\beta_0}}$, the proof of Theorem 3.1 is complete. □

7.7.3. *Proof of Lemma 7.5.* First we write,

(30)

$$\mathbb{E}\left(\sup_{m'} \left((\hat{f}_{m'}(x_0) - f_{m'}(x_0))^2 - \frac{V(m', x_0)}{6} \right)_+ \right) \leq \sum_{m \in \mathcal{M}_n} \mathbb{E}\left(\left((\hat{f}_m(x_0) - f_m(x_0))^2 - \frac{V(m, x_0)}{6} \right)_+ \right).$$

Then

$$\mathbb{E}\left(\left((\hat{f}_m(x_0)-f_m(x_0))^2-\frac{V(m,x_0)}{6}\right)_+\right)\leq\int_0^{+\infty}\mathbb{P}\left(|\hat{f}_m(x_0)-f_m(x_0)|\geq\sqrt{\frac{V(m,x_0)}{6}+u}\right)du.$$

Now we apply the Bernstein inequality (see Birgé and Massart (1998), p.366).

Lemma 7.7. *Let $T_1,\ldots,T_n$ be independent random variables and $S_n(T)=\sum_{i=1}^n[T_i-\mathbb{E}(T_i)]$. Then, for $\eta>0$,*

$$\begin{aligned}P(|S_n(T)-\mathbb{E}(S_n(T))|\geq n\eta) &\leq 2\exp\left(\frac{-n\eta^2/2}{v^2+b\eta}\right)\\ &\leq 2\min\left(\exp\left(-\frac{n\eta^2}{4v^2}\right),\exp\left(-\frac{n\eta}{4b}\right)\right),\end{aligned}\tag{31}$$

where

$$\mathrm{Var}(T_1)\leq v^2 \text{ and } |T_1|\leq b \quad \left(\text{or } \frac{1}{n}\sum_{i=1}^n\mathbb{E}(|T_i|^m)\leq\frac{m!}{2}v^2b^{m-2},\ \forall m\geq 2\right).$$

For our purpose, we have:

$$T_i=\frac{1}{x_0}K_m(\frac{X_i}{x_0}),\quad \eta=\sqrt{\frac{V(m,x_0)}{6}+u},$$

and we need compute v^2 and b. We have $|K_m|\leq\sum_{i=1}^L|\alpha_i|K_m^{(i)}$ where, for all u,

$$K_m^{(i)}(u)\leq\frac{\sqrt{m}}{2\pi}\int_{\mathbb{R}}|(K_m^{(i)})^*|(t/\sqrt{m})|^m dt\leq 2\frac{\sqrt{m}}{\sqrt{2\pi v_i}},$$

by Lemma 2.1. Hence,

$$\frac{1}{x_0}|K_m(\frac{X_i}{x_0})|\leq 2\frac{\sqrt{m}}{x_0}\sum_{i=1}^L|\alpha_i|\frac{1}{\sqrt{2\pi v_i}}:=b.$$

Using the variance bound, we choose, for $C(K)$ given by (17)

$$v^2=C(K)(\|f\|_\infty\vee 1)\frac{\sqrt{m}}{x_0}.$$

We find,

$$\frac{n\eta^2}{4v^2}\geq\frac{\kappa}{24}\log(n)+\frac{nux_0}{4C_1'\sqrt{m}}\ \text{ with } C_1'=C(K)(\|f\|_\infty\vee 1).$$

Now, we use $\sqrt{a+b}\geq\sqrt{a/2}+\sqrt{b/2}$, and denoting by $A=2\sum_{i=1}^L\alpha_i/\sqrt{2\pi v_i}$, we get

$$\frac{n\eta}{4b}\geq\frac{\sqrt{\kappa C_1'}}{8\sqrt{3}A}\sqrt{\log(n)}\frac{\sqrt{nx_0}}{m^{1/4}}+\frac{nx_0}{4\sqrt{2m}A}\sqrt{u}.$$

Thus, as $m\leq(n/\log(n))^2$,

$$n\eta/(4b)\geq\frac{\sqrt{\kappa C_1'x_0}}{8\sqrt{3}A}\log(n)+\log(n)x_0\frac{\sqrt{u}}{4\sqrt{2}A}.$$

Hence, as $x_0 > x_{min}$,

$$\begin{aligned}
&\int_0^{+\infty} P\left(|f_m(x_0) - \hat{f}_m(x_0)| > \sqrt{\frac{V(m,x_0)}{6} + u}\right) du \\
\leq\ & 2\min(n^{-\kappa/24}\int_0^{+\infty} \exp(-\frac{nux_0}{4C_1'\sqrt{m}})du, n^{-\sqrt{\kappa C_1' x_0}/(8\sqrt{3}A)}\int_0^{+\infty} \exp(-\log(n)x_0\frac{\sqrt{u}}{4\sqrt{2}A})du) \\
\leq\ & 2\min(\frac{4C_1'\sqrt{m}}{nx_{min}}n^{-\kappa/24}, \frac{64A^2}{x_{min}^2\log^2(n)}n^{-\sqrt{\kappa C_1' x_{min}}/(8\sqrt{3}A)}) \\
\leq\ & Cn^{-\omega}
\end{aligned}$$

if $\kappa/24 \geq \omega$, $\sqrt{\kappa C_1' x_{min}}/(8\sqrt{3}A) \geq \omega$. Then, we can choose κ large enough such that $\mathrm{card}(\mathcal{M}_n)n^{-\omega} \leq n^{1-\omega} \leq C/n$, that is $\kappa \geq \max(2 \times 24, 12 \times 64A^2/(C_1' x_{min})$.

Hence, the result is complete. □

7.7.4. *Proof of Lemma 7.6.* The proof follows the same line. The new elements are to be taken from (29) in Lemma 7.3 and from Lemma 7.4.

7.8. **Proof of Proposition 4.1.** We have

$$\begin{aligned}
\mathrm{Var}\hat{f}_m(x) &= \frac{1}{nx^2}\left[\int_0^{+\infty} K_m^2(t/x)f(t)dt - \left(\int_0^{+\infty} K_m(t/x)f(t)dt\right)^2\right] \\
&\leq \frac{1}{nx}\int_0^{+\infty} K_m^2(u)f(xu)du.
\end{aligned}$$

Thus,

$$\int_0^{+\infty} x\mathrm{Var}(\hat{f}_m(x))dx \leq \frac{1}{n}\int_0^{+\infty}\int_0^{+\infty} dxduK_m^2(u)f(xu) = \frac{1}{n}\int_0^{+\infty} K_m^2(u)\frac{du}{u}.$$

For $i = 1, \ldots, L$, we have the following uniform bound, using Lemma 2.1:

$$|K_m(u)| \leq \sum_{i=1}^{L} |\alpha_i| K_m^{(i)}(u) \leq \sqrt{m}\sum_{i=1}^{L} |\alpha_i|/\sqrt{2\pi v_i}(1 + o_m(1)),$$

Using that

$$\int_0^{+\infty} K_m^{(i)}(u)\frac{du}{u} = 1 + O(1/m),$$

yields:

$$\int_0^{+\infty} x\mathrm{Var}(\hat{f}_m(x))dx \leq C\frac{\sqrt{m}}{n},$$

□

7.9. **Proof of Proposition 4.2.** We recall the generalized Minkovski inequality. The proof of the following inequality can be found in *e.g.* Tsybakov (2004, p. 161). For all Borel function g on $\mathbb{R} \times \mathbb{R}$, we have

$$\int_{\mathbb{R}}\left(\int_{\mathbb{R}} g(u,x)du\right)^2 dx \leq \left(\int_{\mathbb{R}}\left(\int_{\mathbb{R}} g^2(u,x)dx\right)^{1/2} du\right)^2.$$

As in Proposition 3.2, we start with

$$\mathbb{E}\hat{f}_m(x) - f(x) = x^\ell \int_0^{+\infty} du(u-1)^\ell K_m(u)\int_0^1 \frac{s^{\ell-1}}{(\ell-1)!}\left(f^{(\ell)}(x + sx(u-1))) - f^{(\ell)}(x)\right) ds.$$

Then, as in Tsybakov (2004, Proposition 1.8, p.20), we can apply twice the generalized Minkowski inequality and the assumptions to get the result. For the first application, we set:

$$g(u,x) = 1_{(x>0)}1_{(u>0)}x^\ell(u-1)^\ell K_m(u)I(u,x)$$

with

$$I(u,x) = \int_0^1 \frac{s^{\ell-1}}{(\ell-1)!}\left(f^{(\ell)}(x+sx(u-1))) - f^{(\ell)}(x)\right)ds.$$

We obtain:

$$\int_0^{+\infty}\left(\mathbb{E}\hat{f}_m(x)-f(x)\right)^2 dx \le \left[\int_0^{+\infty} du|u-1|^\ell |K_m(u)|\left(\int_0^{+\infty} x^{2\ell}I^2(u,x)dx\right)^{1/2}\right]^2.$$

For the second application, we set

$$h(s,x,u) = 1_{(x>0)}1_{[0,1]}(s)x^\ell\frac{s^{\ell-1}}{(\ell-1)!}\left(f^{(\ell)}(x+sx(u-1))) - f^{(\ell)}(x)\right).$$

Thus,

$$\int_0^{+\infty}\left(\int_0^1 h(s,x,u)ds\right)^2 dx \le \left[\int_0^1 \frac{s^{\ell-1}}{(\ell-1)!}\left(\int_0^{+\infty} dx\, x^{2\ell}\left(f^{(\ell)}(x+sx(u-1))) - f^{(\ell)}(x)\right)^2\right)^{1/2}\right]^2.$$

By definition of the class $\mathcal{H}_+(\beta,\phi)$, setting $v = 1+s(u-1) = 1-s+us$,

$$\left(\int_0^{+\infty} dx\, x^{2\ell}\left(f^{(\ell)}(vx)-f^{(\ell)}(x)\right)^2\right)^{1/2} \le \frac{\phi(v)}{v^{\ell+(1/2)}}|v-1|^{\beta-\ell} \le \psi(u)|u-1|^{\beta-\ell}.$$

Consequently,

$$\int_0^{+\infty}\left(\int_0^1 h(s,x,u)ds\right)^2 dx \le \psi(u)|u-1|^{\beta-\ell}/\ell!.$$

Finally,

$$\int_0^{+\infty}\left(\mathbb{E}\hat{f}_m(x)-f(x)\right)^2 dx \le \frac{1}{\ell!^2}\left(\int_0^{+\infty}\psi(u)|u-1|^\beta|K_m(u)|du\right)^2 \le \frac{\tilde{C}^2 m^{-\beta}}{\ell!^2}.$$

□

7.10. **Proof of Proposition 4.3.** Set $D(u) = F^2(u)$. Then,

$$D(u) = (1+u^{2\ell+1})\int_0^{+\infty}(f^{(\ell)}(x))^2x^{2\ell}dx - 2u^{2\ell+1}\int_0^{+\infty} f^{(\ell)}(ux)f^{(\ell)}(x)x^{2\ell}dx.$$

By the assumption, D is C^∞ and:

$$\begin{aligned} D'(u) &= (2\ell+1)u^{2\ell}\int_0^{+\infty}(f^{(\ell)}(x))^2x^{2\ell}dx - 2(2\ell+1)u^{2\ell}\int_0^{+\infty} f^{(\ell)}(ux)f^{(\ell)}(x)x^{2\ell}dx \\ &\quad -2u^{2\ell+1}\int_0^{+\infty} f^{(\ell+1)}(ux)f^{(\ell)}(x)x^{2\ell+1}dx \end{aligned}$$

satisfies:

$$\begin{aligned} D'(1) &= -(2\ell+1)\int_0^{+\infty}(f^{(\ell)}(x))^2x^{2\ell}dx - 2\int_0^{+\infty} f^{(\ell+1)}(x)f^{(\ell)}(x)x^{2\ell+1}dx \\ &= -\left[x^{2\ell+1}(f^{(\ell)}(x))^2\right]_0^{+\infty} = 0. \end{aligned}$$

Next,

$$\begin{aligned} D''(u) &= 2\ell(2\ell+1)u^{2\ell-1}\int_0^{+\infty}(f^{(\ell)}(x))^2x^{2\ell}dx - 2(2\ell)(2\ell+1)u^{2\ell-1}\int_0^{+\infty}f^{(\ell)}(ux)f^{(\ell)}(x)x^{2\ell}dx \\ &\quad -4(2\ell+1)u^{2\ell}\int_0^{+\infty}f^{(\ell+1)}(ux)f^{(\ell)}(x)x^{2\ell+1}dx - 2u^{2\ell+1}\int_0^{+\infty}f^{(\ell+2)}(ux)f^{(\ell)}(x)x^{2\ell+2}dx \end{aligned}$$

satisfies

$$\begin{aligned} D''(1) &= -2\ell(2\ell+1)\int_0^{+\infty}(f^{(\ell)}(x))^2x^{2\ell}dx - 4(2\ell+1)\int_0^{+\infty}f^{(\ell+1)}(x)f^{(\ell)}(x)x^{2\ell+1}dx \\ &\quad -2\int_0^{+\infty}f^{(\ell+2)}(x)f^{(\ell)}(x)x^{2\ell+2}dx. \end{aligned}$$

Using $F'(1)=0$ and another integration by parts yields

$$D''(1) = 2\int_0^{+\infty}(f^{(\ell+1)}(x))^2x^{2\ell+2}dx \neq 0.$$

Using a Taylor expansion,

$$F(u) = \frac{1}{2}(u-1)^2\int_0^1 F''(1+s(u-1)ds,$$

where $|F''(u)| \leq c(1+u^{2\ell+1})$. Hence, the result.□

7.11. **Checking Assumption (20) on examples.** It is enough to check it when K_m is the density of $\bar{U}_m = (U_1+\ldots+U_m)/m$ with $U_1,\ldots,U_m$ i.i.d. with density K satisfying (H1)-(H3). To get (20), we can prove that the sequence

$$(m^{\beta/2}\psi(\bar{U}_m))|\bar{U}_m-1|^\beta) \tag{32}$$

is uniformly integrable. This holds if, for some $\delta>1$,

$$\sup_m \mathbb{E}\left[m^{\beta/2}\psi(\bar{U}_m))|\bar{U}_m-1|^\beta\right]^\delta.$$

For $f(x)=1_{[0,1]}(x)$, we have $\phi(u)=1$, $\ell=0,\beta=1/2$. Thus, we must check that

$$m^{1/4}\frac{|\bar{U}_m-1|^{1/2}}{\bar{U}_m^{1/2}}, m\geq 1,$$

is uniformly integrable. For this, we compute, for K equal to the density $G(a,a)$ with $m>2/a$ and $\delta=4$:

$$m\mathbb{E}\frac{(\bar{U}_m-1)^2}{\bar{U}_m^2} = \frac{m(2+am)}{(am-1)(am-2)} = \frac{1}{a}(1+O(1)).$$

Hence, condition (20) holds.
For $f^{(\ell)}(x)=(-1)^\ell e^{-x}1_{x>0}$, we have $\beta=l+1$, $\phi(u)\leq c(1+u^{2\ell})$. We take $\delta=2$ and check that there exists m_0 such that

$$\sup_{m\geq m_0} m^{\ell+1}\mathbb{E}(\frac{1}{\bar{U}_m^{\ell+(1/2)}}+\bar{U}_m^{\ell-(1/2)})(\bar{U}_m-1)^{2\ell+2} < +\infty. \tag{33}$$

Using the property of moments (see Lemma 7.1) and the Cauchy-Schwarz inequality, it is enough to check that

$$\mathbb{E}(\frac{1}{\bar{U}_m^{2(2\ell+1)}}) = 1+O(1/m).$$

For K equal to the density $G(a,a)$, this holds for $am > 2(2\ell+1)$.

7.12. **Proof of Theorem 4.2.**

7.12.1. *A preliminary Lemma.* The Lemmas of Section 7.7.1 are also used here, and must be completed by the following one.

Lemma 7.8.

$$\int_0^{+\infty} Var\hat{f}_{m,m'}(x)\, x\, dx \leq \frac{\sqrt{m\wedge m'}}{n} C(K)(1+o(1)),$$

where the constant $C(K)$ (see (17)) does not depend on the density f.

Proof. For the second bound, by integrating, we get:

$$\int_0^{+\infty} \mathrm{Var}\hat{f}_{m,m'}(x)\, x\, dx \leq \frac{1}{n}\int_0^{+\infty} (K_m \odot K_{m'}(u))^2\, du/u.$$

To conclude, we use Lemma 7.3 again and the fact that:

$$\int_0^{+\infty} |K_m \odot K_{m'}(u)| du/u \leq \sum_{i,j} |\alpha_i\alpha_j| \mathbb{E}(1/\bar{U}_m^{(i)})\mathbb{E}(1/\bar{V}_{m'}^{(j)}) = 1 + O(1/m) + O(1/m'). \tag{34}$$

□

7.12.2. *Proof of Theorem 4.2.* Recall that $w(x) = x \wedge 1$ and that we have set $< f,g >_w = \int_0^{+\infty} f(x)f(x)w(x)dx$, $||f||_w = < f,f >_w^{1/2}$. First note that the definition of $\hat{m}$ implies that $A(\hat{m}) + V(\hat{m}) \leq A(m) + V(m)$ for all $m \in \mathcal{M}_n$. Hence, for m any element of $\mathcal{M}_n$, we can write the decomposition

$$\begin{aligned}
\|\tilde{f} - f\|_w^2 &\leq 3(\|\hat{f}_{\hat{m}} - \hat{f}_{m,\hat{m}}\|_w^2 + \|\hat{f}_{m,\hat{m}} - \hat{f}_m\|_w^2 + \|\hat{f}_m - f\|_w^2) \\
&\leq 3(A(m) + V(\hat{m})) + 3(A(\hat{m}) + V(m)) + 3\|\hat{f}_m - f\|_w^2 \\
&\leq 6(A(m) + V(m)) + 3\|\hat{f}_m - f\|_w^2.
\end{aligned}$$

Therefore,

$$\mathbb{E}(\|\tilde{f} - f\|_w^2) \leq 3\mathbb{E}(\|\hat{f}_m - f\|_w^2) + 6V(m) + 6\mathbb{E}(A(m)).$$

Let us study $A(m)$ (see (21)). Let $\mathbb{E}(\hat{f}_m(x)) = f_m(x)$ and $\mathbb{E}(\hat{f}_{m,m'}(x)) = f_{m,m'}(x)$. Then

$$\|\hat{f}_{m'} - \hat{f}_{m,m'}\|_w^2 \leq 3\|\hat{f}_{m'} - f_{m'}\|_w^2 + 3\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 + 3\|f_{m'} - f_{m,m'}\|_w^2.$$

By Lemma 7.2, for all $m, m' \in \mathcal{M}_n$,

$$\|f_{m'} - f_{m,m'}\|_w^2 = \int \left(\int_0^{+\infty} B_m f(xu) K_{m'}(u) du\right)^2 w(x)dx.$$

Therefore, as $w(x) \leq 1$, setting $c = \sum_{i=1}^L |\alpha_i|$, and using that each $K_{m'}^{(i)}$ is a density, we obtain:

$$\begin{aligned}
\|f_{m'} - f_{m,m'}\|_w^2 &\leq c\int_0^{+\infty} \sum_{i=1}^L |\alpha_i| \left(\int_0^{+\infty} (B_m f)(xu) K_{m'}^{(i)}(u) du\right)^2 dx \\
&\leq c\sum_{i=1}^L |\alpha_i| \int\int (B_m f)^2(xu) K_{m'}^{(i)}(u) du dx \\
&\leq c\int_0^{+\infty} (B_m f)^2(v) dv \sum_{i=1}^L |\alpha_i| \int_0^{+\infty} \frac{K_{m'}^{(i)}(u)}{u} du,
\end{aligned}$$

having used Fubini and the change of variable $v = xu$. Now, $\int K_{m'}^{(i)}(u)/u \; du = C_{m'}^{(i)} = 1 + O(1/m') \leq 2$. Therefore

$$\begin{aligned} A(m) &\leq 3\sup_{m'}\left(\|\hat{f}_{m'} - f_{m'}\|_w^2 - \frac{V(m')}{6}\right)_+ + 3\sup_{m'}\left(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 - \frac{V(m')}{6}\right)_+ \\ &\quad +3\sup_{m'}\|f_{m'} - f_{m,m'}\|_w^2 \\ &\leq 3\sup_{m'}\left(\|\hat{f}_{m'} - f_{m'}\|_w^2 - \frac{V(m')}{6}\right)_+ + 3\sup_{m'}\left(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 - \frac{V(m')}{6}\right)_+ \\ &\quad +6c^2\int (B_m f)^2(v)dv. \end{aligned}$$

Now, we can prove the following Lemmas:

Lemma 7.9. *Under the assumptions of Theorem 4.2, we have*

$$\mathbb{E}\left(\sup_{m'}\left(\|\hat{f}_{m'} - f_{m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right) \leq \frac{C}{n}.$$

Lemma 7.10. *Under the assumptions of Theorem 4.2, we have*

$$\mathbb{E}\left(\sup_{m'}\left(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right) \leq \frac{C}{n}.$$

This yields that, $\forall m \in \mathcal{M}_n$,

$$\mathbb{E}(\|\tilde{f} - f\|_w^2) \leq 3\mathbb{E}(\|\hat{f}_m - f\|_w^2) + 6V(m) + 6c^2\int (B_m f)^2(v)dv + \frac{6C}{n}.$$

As $\mathbb{E}(\|\hat{f}_m - f\|_w^2) \lesssim (V(m) + \int (B_m f)^2(v)dv)$, the proof of Theorem 4.2 is complete. □

7.12.3. *Proof of Lemma 7.9.* First we write,

$$\mathbb{E}\left(\sup_{m'}\left(\|\hat{f}_{m'} - f_{m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right) \leq \sum_{m\in\mathcal{M}_n}\mathbb{E}\left(\left(\|\hat{f}_m - f_m\|_w^2 - \frac{V(m)}{6}\right)_+\right). \tag{35}$$

Next, we note that $\|\hat{f}_m - f_m\|_w^2 = \sup_{t,\|t\|_w=1}\langle \hat{f}_m - f_m, t\rangle_w^2$, and the supremum can be taken over a dense countable family of functions t such that $\|t\|_w = 1$; we denote by $\mathcal{B}(1)$ this set. Thus,

$$\nu_n(t) = \langle \hat{f}_m - f_m, t\rangle_w^2 = \frac{1}{n}\sum_{j=1}^n \int \frac{1}{x}\left[K_m(\frac{X_j}{x}) - \mathbb{E}K_m(\frac{X_j}{x})\right] t(x)w(x)dx$$

is a centered empirical process, and we can apply the Talagrand inequality (see Talagrand (1996)):

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}(1)}(\nu_n^2(t) - 4H^2)_+\right] \leq C_1\left[\frac{b}{n}e^{-C_2\frac{nH^2}{b}} + \frac{M_1^2}{n^2}e^{-C_3\frac{nH}{M_1}}\right] \tag{36}$$

where C_1, C_2, C_3 are three numerical constants and H, b, M_1 are defined by

$$\mathbb{E}(\sup_{t\in\mathcal{B}(1)}\nu_n^2(t)) \leq H^2, \quad \sup_{t\in\mathcal{B}(1)}\mathrm{Var}\left(\int \frac{1}{x}K_m(\frac{X_1}{x})t(x)w(x)dx\right) \leq b$$

and

$$\sup_{t\in\mathcal{B}(1)}\sup_u\left|\int \frac{1}{x}K_m(\frac{u}{x})t(x)w(x)dx\right| \leq M_1.$$

It follows from the definition of ν_n that

$$\mathbb{E}(\sup_{t\in\mathcal{B}(1)} \nu_n^2(t)) \leq \mathbb{E}(\|\hat{f}_m - f_m\|_w^2) \leq 2C(K)\sqrt{m}/n := H^2$$

where $C(K)$ is defined in (17).

Next, for $\|t\|_w = 1$, we have

$$\begin{aligned}\left|\int \frac{1}{x}K_m(\frac{u}{x})t(x)w(x)dx\right| &\leq \left(\int (\frac{1}{x}K_m(\frac{u}{x}))^2 w(x)dx \int t^2(x)w(x)dx\right)^{1/2} \\ &= \left(\int (\frac{v}{u}K_m(v))^2 w(\frac{u}{v})\frac{u}{v^2}dv\right)^{1/2} \\ &= \left(\int (K_m(v))^2[w(\frac{u}{v})\frac{v}{u}]\frac{dv}{v}\right)^{1/2}\end{aligned}$$

and by noting that $w(x)/x = w(1/x) \leq 1$, we get

$$\left|\int \frac{1}{x}K_m(\frac{u}{x})t(x)w(x)dx\right| \leq \left(\int (K_m(v))^2\frac{dv}{v}\right)^{1/2} \leq \|K_m\|_\infty^{1/2}(\int |K_m(v)|\ dv/v)^{1/2}.$$

Again, $I_m = \int_0^{+\infty} |K_m(v)|/v dv \leq 2\sum_{i=1}^L |\alpha_i|$ and $\|K_m\|_\infty \leq 2\sqrt{m}\sum_{i=1}^L |\alpha_i|/\sqrt{2\pi v_i}$. Hence, we can take $M = C(K)m^{1/4}$. Now, we study

$$\begin{aligned}\mathbb{V} &:= \mathrm{Var}\left(\int \frac{1}{x}K_m(\frac{X_1}{x})t(x)w(x)dx\right) \leq \mathbb{E}\left(\int_0^{+\infty}(1/x)K_m(X_1/x)t(x)w(x)dx\right)^2 \\ &\leq \int_{(0,+\infty)^3} f(\xi)(1/x)K_m(\xi/x)t(x)w(x)(1/y)K_m(\xi/y)t(y)w(y)dxdyd\xi.\end{aligned}$$

First, $\int f(\xi)K_m(\xi/x)K_m(\xi/y)d\xi = x\ \int K_m(u)K_m(xu/y)f(xu)du$. Hence,

$$\mathbb{V} \leq \int_0^{+\infty} K_m(u)du\left(\int_0^{+\infty} t(x)w(x)f(xu)(\int_0^{+\infty} t(y)w(y)K_m(xu/y)(1/y)dy)dx\right). \tag{37}$$

Next,

$$\int_0^{+\infty} t(y)w(y)K_m(xu/y)(1/y)dy \leq \left[\int_0^{+\infty} t^2(y)w(y)dy \int_0^{+\infty}(K_m(xu/y)(1/y))^2 w(y)dy\right]^{1/2},$$

and (with $v = xu/y$),

$$\int_0^{+\infty}(K_m(xu/y)(1/y))^2 w(y)dy \leq \frac{1}{xu}\int_0^{+\infty} K_m^2(v)dv \leq \frac{1}{xu}C\sqrt{m}.$$

This yields:

$$\mathbb{V} \leq \int_0^{+\infty} K_m(u)\frac{1}{\sqrt{u}}du\left(\int_0^{+\infty} t(x)w(x)f(xu)\frac{1}{\sqrt{x}}dx\right)C^{1/2}m^{1/4}.$$

Then, using again $w(x)/x = w(1/x) \leq 1$,

$$|\int_0^{+\infty} t(x)w(x)f(xu)\frac{1}{\sqrt{x}}dx| \leq \left[\int t^2(x)\frac{w^2(x)}{x}dx\int f^2(xu)dx\right]^{1/2} \leq \frac{1}{\sqrt{u}}||t||_w||f||.$$

Finally,

$$\mathbb{V} \leq C^{1/2}m^{1/4}\int_0^{+\infty} K_m(u)\frac{1}{u}du||t||_w||f|| \leq C^{1/2}m^{1/4}||f||(1+O(1)).$$

Thus, we can take $b = C'm^{1/4}$. Lastly,

$$\frac{nH}{M} \propto n^{1/2}, \quad \frac{nH^2}{b} \propto m^{1/4}.$$

This yields

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}(1)} (\nu_n^2(t) - 4H^2)_+\right] \leq C_1' \left[\frac{m^{1/4}}{n} e^{-C_2' m^{1/4}} + \frac{\sqrt{m}}{n^2} e^{-C_3' n^{1/2}}\right].$$

As $m \leq n^2$ in $\mathcal{M}_n$,

$$\mathbb{E}\left[\sup_{t\in\mathcal{B}(1)} (\nu_n^2(t) - 4H^2)_+\right] \leq \frac{C_4}{n} m^{1/4} e^{-C_5 m^{1/4}}.$$

Now, reminding of (35), we get

$$\mathbb{E}\left(\sup_{m'}\left(\|\hat{f}_{m'} - f_{m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right) \leq \frac{C_4}{n} \sum_{m\in\mathcal{M}_n} m^{1/4} e^{-C_5 m^{1/4}} \leq \frac{C_6}{n}$$

This ends the proof of Lemma 7.9. □.

7.12.4. *Proof of Lemma 7.10.* The proof of Lemma (7.10) follows the same line as previously with K_m replaced by $K_m \odot K_{m'}$, where m is fixed and the sum is now taken over m' in $\mathcal{M}_n$.

(38)

$$\mathbb{E}\left(\sup_{m'}\left(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right) \leq \sum_{m'\in\mathcal{M}_n} \mathbb{E}\left(\left(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 - \frac{V(m')}{6}\right)_+\right).$$

Thus, we apply Inequality (36) to the empirical process

$$\nu_n^*(t) = \langle \hat{f}_{m,m'} - f_{m,m'}, t\rangle_w^2 = \frac{1}{n}\sum_{j=1}^n \int \frac{1}{x}\left[K_m \odot K_{m'}(\frac{X_j}{x}) - \mathbb{E}K_m \odot K_{m'}\frac{X_j}{x}\right) t(x)w(x)dx$$

where we have to find

$$\mathbb{E}(\sup_{t\in\mathcal{B}(1)} (\nu_n^*)^2(t)) \leq H_*^2, \ \sup_{t\in\mathcal{B}(1)} \mathrm{Var}\left(\int \frac{1}{x} K_m \odot K_{m'}(\frac{X_1}{x}) t(x)w(x)dx\right) \leq b^*$$

and

$$\sup_{t\in\mathcal{B}(1)} \sup_u \left|\int \frac{1}{x} K_m \odot K_{m'}(\frac{u}{x}) t(x)w(x)dx\right| \leq M_1^*.$$

For the first term, we can apply Lemma 7.4 to get

$$\mathbb{E}(\sup_{t\in\mathcal{B}(1)} (\nu_n^*)^2(t)) \leq \mathbb{E}(\|\hat{f}_{m,m'} - f_{m,m'}\|_w^2 \leq C(K)\frac{\sqrt{m\wedge m'}}{n} \leq C(K)\frac{\sqrt{m'}}{n} := H_*^2.$$

Next, by analogy with the proof of Lemma 7.9, we have

$$\left|\int \frac{1}{x} K_m \odot K_{m'}(\frac{u}{x}) t(x)w(x)dx\right| \leq \|K_m \odot K_{m'}\|_\infty \int |K_m \odot K_{m'}(v)|/v dv.$$

This gives, by using Lemma 7.3 and inequality (34)

$$\left|\int \frac{1}{x} K_m \odot K_{m'}(\frac{u}{x}) t(x)w(x)dx\right| \leq C(m\wedge m')^{1/4} \leq C(m')^{1/4} := M_1^*.$$

In the same way, we get $b^* = C'm^{1/4}$. The bounds being the same as for Lemma 7.9, the conclusion is also analogous. □

References

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* National Bureau of Standards Applied Mathematics Series, 55 For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2**, no. 1, 4168.

Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.***63**, no. 2, 167-241.

Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, **113**, no. 3, 301-413.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.

Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.* **5**, 131.

evroye, L. (1989). The double kernel method in density estimation. *Ann. Inst. H. Poincar Probab. Statist.* **25**, no. 4, 533580.

Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimates. *Ann. Statist.* **24**, no. 6, 2499-2512.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist. 24*, no. 2, 508-539.

Feller W. (1971). *An introduction to probability theory and its applications, Volume II, second edition*, Wiley.

Goldenshluger A. and Lepski, O. (2010). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality (*preprint*) ArXiv:1009.1016.

Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, Inc. Harcourt Brace Jovanovich, Publishers, New York-London.

Juditsky, A. (1997). Wavelet estimators: adapting to unknown smoothness. Math. Methods Statist. 6, no. 1, 125.

Kendall, M. and Stuart, A.O. (1987). *J. Keith Kendall's advanced theory of statistics.* Vol. 1. Distribution theory. Fifth edition. The Clarendon Press, Oxford University Press, New York.

Kerkyacharian, G., Lepski, O. and Picard, D. (2001). Non-linear estimation in anisotropic multi-index denoising. Probab. Theory Related Fields **121**, 137-170.

Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli* **2**, no. 3, 229-247.

Ledoux, M. (1997). On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, **1**, 63-87.

Massart, P. (2007). *Concentration inequalities and model selection.* Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 623, 2003. Lecture Notes in Mathematics, 1896. Springer, Berlin.

Sato, K. (1999). *Lévy processes and infinitely divisible distributions.* Cambridge Studies in Advanced Mathematics, 68. Cambridge University Press, Cambridge.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, **8**, no. 6, 1348-1360.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505-563.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.* Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer Series in Statistics. Springer, New York.