CONVOLUTION POWER KERNELS FOR DENSITY ESTIMATION

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

Abstract. We propose a new type of non parametric density estimators fitted to random variables with lower or upper-bounded support. To illustrate the method, we focus on 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 exponent m plays the role of the bandwidth. We study the pointwise Mean Square Error and propose a pointwise adaptive estimator. The risk of the adaptive estimator satisfies an almost oracle inequality. A noteworthy result is that the adaptive rate is in correspondence with the smoothness properties of the unknown density as a function on $(0, +\infty)$. The adaptive estimators are illustrated on simulated data. We compare our approach with the classical kernel estimators.

January 3, 2012

Keywords. Adaptive estimators. Density estimation. Infinitely divisible distributions. Kernel estimators. Lower bounded support.

1. Introduction

Let $X_1, \dots, X_n$ be n *i.i.d.* 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 several 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), Efromovich (1999). See also Butucea (2001) for a detailed discussion.

In the case of upper or lower-bounded observations, even if f is very smooth as a function on its support, f may be not even continuous as a function on $(-\infty, +\infty)$. For simplicity we consider here the case of density estimation for nonnegative random variables where f is zero on $\mathbb{R}^-$. Our aim is to propose estimators free of boundary effects near zero.

The behavior of kernel estimators has been described at length, see e.g. Silverman (1986) or Wand and Jones (1995), and in particular the difference between interior points of the support of the density under estimation and boundary points: specifically, it is well known that the bias is larger at boundary points than inside the support. To remove the so-called "boundary effects", many improvements of kernel density estimators have been proposed, as reflection method (see Cline and Hart, 1991), boundary kernel method (Gasser *et al.* (1985), Jones (1993), Zhang and Karunamuni (2000)), the pseudo-data method (Cowling and Hall (1996)), the transformation method (Wand *et. al.* (1991), Marron and Ruppert (1994)), the local linear method (Cheng (1997), Cheng *et al.* (1997), Zhang and Karunamuni (1998)) and also other methods as in Zhang *et al.* (1999) or Hall and Park (2002). Generalizations of all these methods have

[(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.

been described by Karunamuni and Alberts (2005,2006). Most of these methods are devoted to specific behavior of the density near of the boundary points (e.g. the reflection method is specially designed for the case the derivative of f at this point is zero) or have been designed under the assumption that f admits two bounded derivatives. Moreover, risk bounds are provided for fixed bandwidths and the random selected bandwidth is often studied via simulations only. We propose here another method, including bandwidth selection and for which the risk bound is provided even for the estimator with data driven selected bandwidth; our method can also adapt to any order of regularity of the density under estimation. The bandwidth is selected at each point of estimation with a simple criterion instead of being selected globally after the computation of a much more complicated quantity. Note that allowing the bandwidth to be different at different locations is not a recent strategy (see Wand *et al.* (1991), Mack and Rosenblatt (1979), Hall (1983) and more recently, applying the so-called Lepski-method to density estimation with local bandwidth selection, Butucea (2001) and the references therein). Lastly, the estimator is not defined at 0 (i.e. at the boundary point) so that it is successful even when densities tend to infinity at this point.

To realize all these performances, we introduce a new type of kernel estimators, based on kernels which are approximations of the Dirac mass 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.

The basic idea of their construction is the following. Consider a density K on $(0, +\infty)$, with expectation 1, and let $U_1, \dots, U_m$ 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

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

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. In Section 2, we precise the assumptions on the densities K used to build estimators, extend the definition of (2) and give examples of densities K for which K_m is explicit. Section 3 is devoted to pointwise estimation. For this, we show that the extended kernels allow us to obtain the best possible order for the bias term, fitted with the smoothness of the unknown density as a function on $(0, +\infty)$. We also define a new notion of kernel of order ℓ. The case where K is an infinitely divisible density allows to give an explicit construction of this kernel of order ℓ. This construction is closely related to the construction of Kerkyacharian *et al.* (2001). Proposition 3.3 gives the pointwise mean squared error: it shows that the convolution exponent m plays the role of the bandwidth and has to be adjusted for risk minimization. In Section 3.3, we study the data driven selection of m. For this, 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. These iterated kernels are involved in the definition of the selection criterion. We prove in Theorem 3.1 that the adaptive estimator reaches the same optimal adaptive rate as the one obtained by Butucea (2001). In Section 4, numerical simulation results are provided to illustrate the estimation method. Moreover, we consider the kernel estimators described in Goldenschluger and Lepski (2011) and implement a local adaptive bandwidth selection which is new. Both methods are convincing and we discuss their respective advantages and drawbacks. For a comparison purpose, we also implement the standard kernel estimator available in Matlab. A quantitative measure of the performances of estimators is provided based on risk computations, which shows that our estimator detects well the irregularity point. Section 5 contains concluding remarks. Proofs are gathered in Section 6.

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 Lemma 6.1 in Appendix.

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 $m(1-\bar{U}_m)^2/\bar{U}_m$ 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 6.2).

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.

2.3.1. *Uniform kernels and splines.* A first natural case to consider in the uniform kernel $K(x) = (1/2)\mathbf{1}_{[0,2]}(x)$, corresponding to the uniform distribution on $[0, 2]$. Then Formula (9) in Killmann and von Collani (2001) (see also Rényi (1970)) yields

$$K_m(x) = \frac{m}{(m-1)!2^m}\sum_{i=0}^{\tilde{m}(m,x)} (-1)^i \binom{m}{i} (mx - 2i)^{m-1} \quad \text{if } 0 \leq x \leq 2$$

and 0 otherwise, and $\tilde{m}(m, x) = \lceil mx/2 \rceil$ is the largest integer less than $mx/2$. Here, clearly, K is not continuous on $(0, +\infty)$ and successive convolutions increase the regularity. Therefore, the exponent m plays here clearly the role of regularity parameter and a selection procedure on m is in fact a local regularity estimation.

2.3.2. *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

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

2.3.3. *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)

$$\tilde{K}_\nu(u) = \tilde{K}_{-\nu}(u) = \frac{\sqrt{\pi}(z/2)^\nu}{\Gamma(\nu + 1/2)} \int_0^{+\infty} e^{-z\cosh(t)} (\sinh t)^{2\nu} dt. \tag{7}$$

We use the notation $\tilde{K}_\nu$ for the Bessel function to avoid confusion with the density K_m.
We obtain $e^z \tilde{K}_{3/2}(z) = \sqrt{\pi/2}(z^{-1/2} + z^{-3/2})$, after the successive changes of variables $e^t = u$ and $s = u + u^{-1} - 2$. 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), $\mathbb{E}(\hat{f}_m(x) - f(x))^2$, 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* $1 \le k \le \ell$ *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. Note that a single kernel with $L = 1$ is of order one.
The Hölder-type class on the interval $\mathbf{T} \subset \mathbb{R}$ with constant S and regularity β is defined by:

$$\begin{aligned}\Sigma(\mathbf{T}, \beta, S) \;=\; & \{f : \mathbf{T} \to \mathbb{R}, \\ & f^{(\ell)}\text{exists for } \ell = \lfloor \beta \rfloor \text{and } |f^{(\ell)}(x) - f^{(\ell)}(x')| \le S|x-x'|^{\beta-\ell}, \text{for all } x, x' \in \mathbf{T}\}.\end{aligned}$$

We obtain the following bias bound.

Proposition 3.2. *Assume that* f *belongs to the class* $\Sigma((0,+\infty), \beta, S)$*. 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 Definition 3.1. Then,*

$$|\mathbb{E}\hat{f}_m(x) - f(x)| \le \frac{S\, C_0 x^\beta}{\ell!} \frac{1}{m^{\beta/2}} \tag{8}$$

where C_0 *depends on* β*, the coefficients* α_j *and the moments of the* $K^{(j)}$*'s.*

Remark 3.1. *Note that the regularity condition for* f *is on* $\mathbf{T} = (0,+\infty)$*. Thus, exponential and Gamma densities which are* C^∞ *on* $(0,+\infty)$ *have a very low bias. On the contrary, standard kernel estimation requires a symmetric interval* $\mathbf{T}$*, so that the same densities will have a small value for* β *near* $x = 0$*, implying high bias.*

Remark 3.2. *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* $\Sigma((0,+\infty), \beta, S)$*. 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,*

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

where

$$C_1 = \frac{(1+o_m(1))}{\sqrt{2\pi}} ||f||_\infty \left(\sum_{1 \le i,j \le L} \frac{\alpha_i \alpha_j}{\sqrt{v_i + v_j}} \right), \quad C_2 = S \frac{C_0}{\lfloor \beta \rfloor !}, \tag{9}$$

where $o_m(1)$ *does not depend on* x *and* C_0 *is the constant appearing in (8).*

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 with regularity parameter β (see Stone (1980)).

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^*,

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

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 developing $it + \psi(t)$ 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

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

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 3.3. • *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 of Gamma and inverse Gaussian kernels fit in this framework. The density Gamma $G(a_0, a_0)$ has Lévy density $n(x) = x^{-1}a_0 e^{-a_0 x} 1_{x>0}$. The Inverse Gaussian density $IG(a_0, a_0)$ has Lévy density $n(x) = a_0(2\pi x^3)^{-1/2}e^{-a_0^2 x/2}1_{x>0}$.*

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 $\Sigma((0, +\infty), \beta_0, S)$ where the true value $\beta_0 > 0$ is unknown. We consider kernels $K = \sum_{i=1}^{L} \alpha_i K^{(i)}$ of order $l \geq \beta_0$, that is kernels with orders large enough.

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:

$$\mathbb{E}\left(\hat{f}_{m_{opt}(x_0)}(x_0) - f(x_0)\right)^2 \leq (C_1 + C_2^2)n^{-2\beta_0/(2\beta_0+1)},$$

where C_1 and $C_2 = C_2(\beta_0)$ are defined in (9). 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 $m = k^2x_0^2$. 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

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

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

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

where κ is a numerical constant and

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

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

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

We can prove the following result.

Theorem 3.1. *Assume that f belongs to the class $\Sigma((0, +\infty), \beta_0, S)$ with $\beta_0 > 0$ unknown. Let K be a kernel of order ℓ in the sense of Definition 3.1 with $\ell \geq \lfloor \beta_0 \rfloor$. Consider the estimator $\tilde{f}(x_0) = \hat{f}_{\hat{m}(x_0)}(x_0)$ with $x_0 \geq x_{min} > 0$. 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 $x_n \leq C y_n$ 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 4. We also rewrite a local version of the method of Goldenschluger and Lepski (2011), 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:*

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

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

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

Proof of Proposition 3.6 Recall that $K_m^{(i)} \odot K_{m'}^{(j)}(u) = \int_0^{+\infty} K_m^{(i)}(\frac{u}{v}) K_{m'}^{(j)}(v) dv/v$. 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 $\delta_{ij}^2 = a_i^2 m u + +a_j^2 m'$, $\quad \gamma_{ij}^2 = a_i^2 m/u + +a_j^2 m'$. 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.$\square$

The Bessel function $\tilde{K}_0$ is available in Matlab. Various expressions are given in Abramowitz and Stegun (1964).

4. Practical implementation.

4.1. **Standard kernel estimators with local adaptive bandwidth selection.** We compare our method with the one, based on iterated kernels, described in Goldenschluger and Lepski (2011), 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 $\Sigma(\mathbb{R}, \beta, S)$. 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 $k = 1, \dots, \ell$, 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 (2011), only the global bandwidth selection is studied. Let us describe how local automatic bandwidth selection can be performed. We define

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

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

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

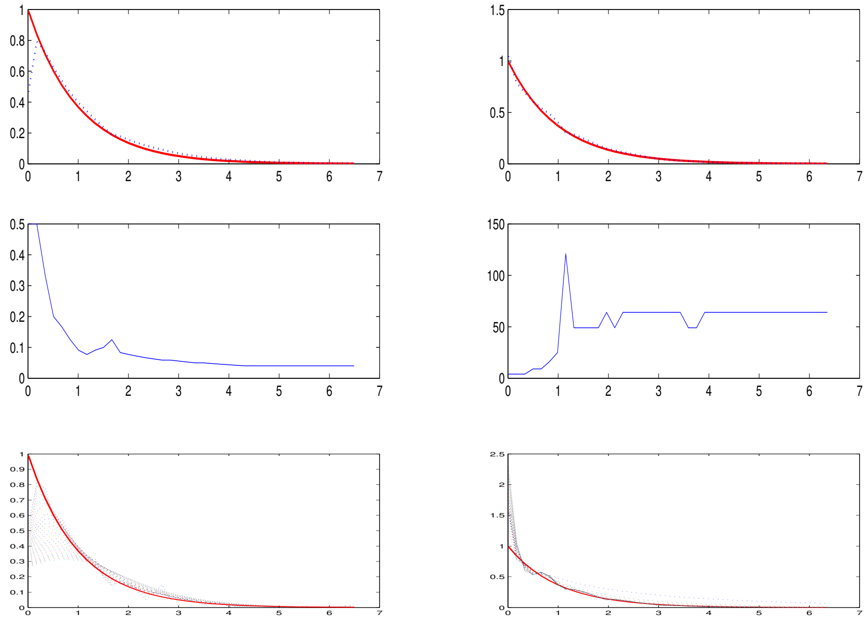

FIGURE 1. Exponential $\mathcal{E}(1)$ density. True function (bold) and estimation for one path (dotted) with $n = 1000$, GL method (up left), our method (up right), selected h (middle left) or m (middle right), true function (bold) and set of all the estimators proposed to the selection algorithm (bottom).

and

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

If f is bounded and in a Hölder class $\Sigma(\mathbb{R}, \beta, S)$ 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. This estimator is abbreviated as "GL"-method. In practice, Gaussian kernels are used to ease the computation of $K_h \star K_{h'}$.

We also compare our method with the kernel estimator of the function `ksdensity` of Matlab, which provides a kernel density estimator with automatically selected bandwidth.

4.2. **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 $\mathcal{H}_n = \{h_k = k/M, k = 1, \ldots, M\}$ and $\mathcal{M}_n = \{m_k = k^2, k = 1, \ldots, M\}$ with $M = 25$. The function is estimated on $[a, b]$ with $a = \max(\min(X), 0.01 \max(X))$ and $b = \max(X)$ where $X = (X_1, \ldots, X_n)$ 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 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

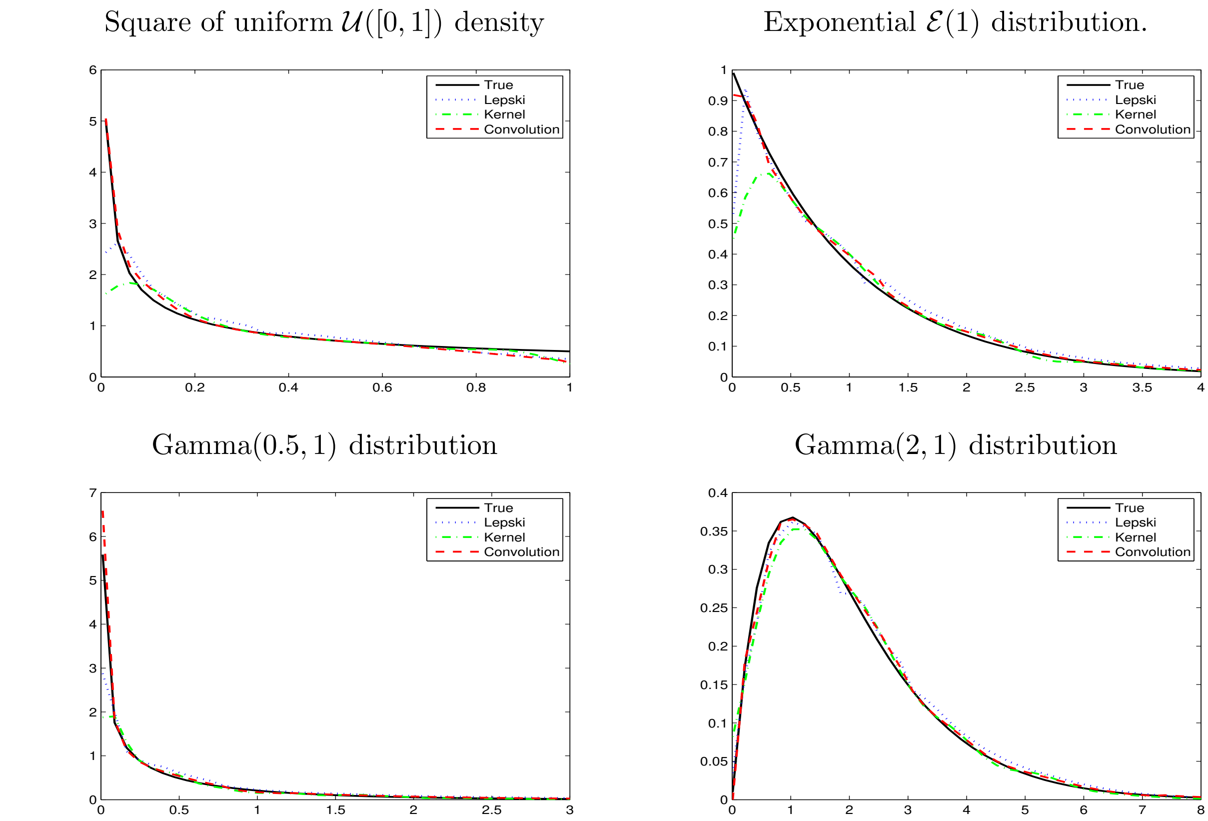

FIGURE 2. True density (bold) and estimated densities with the three methods: standard kernel, GL method, our method ("convolution"), $n = 1000$.

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 show in Figure 2 the difference between the three methods for four different functions. For the distribution which is continuous near zero, the Gamma$(2, 1)$ density, the three methods are obviously equivalent. Then, the case of $0 < f(0) < +\infty$ is illustrated with an exponential $\mathcal{E}(1)$ density, and the case where $f(0) = +\infty$ with the Gamma$(0.5, 1)$ and the square of Uniform variable density: here we can see that the peaks are better estimated by our estimator. We can see that the Gamma$(2, 1)$, which has no discontinuity near 0, is perfectly estimated by the three methods. Now, it is also clear that a classical kernel misses the peaks of the three other distributions, while the GL-method is slightly better thanks to the point-by-point bandwidth choice (the kernel is the same as for the matlab `ksdensity` function), and our methods gives better account of the type of peak, even if it tends a little to over-estimate it.

To assert these observations, we have computed mean squared errors over $Q = 400$ repetitions at points near of the origin for the three methods and the following distributions: the three Gamma distributions ($\Gamma(0.5, 1)$, $\Gamma(1, 1)$ also called exponential $\mathcal{E}(1)$ distribution, $\Gamma(2, 1)$; the square of the uniform variable distribution; the absolute value of a Gaussian; a generalized Pareto distribution $\mathcal{G}P(4, 1, 0)$ where the $\mathcal{G}P(k, \sigma, \theta)$ has distribution $\sigma^{-1}(1+k\sigma^{-1}(x-\theta))^{-1-1/k}$ for $x > \theta$ and $k > 0$. More precisely, Table 4.2 gives the square-root of $\sum_{u=1}^{Q}(\hat{f}_{(u)}(x) - f(x))^2$

	$x_0 =$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	1
$\Gamma(0.5,1)$	Kernel	18.69	9.94	5.95	3.55	1.95	0.85	0.08	0.45	0.80	1.01
	GL	14.09	5.11	1.32	0.47	1.18	1.30	1.10	0.80	0.51	0.31
	Convolution	1.75	3.63	2.05	0.43	0.31	0.75	0.87	1.05	1.30	1.33
$\mathcal{E}(1)$	Kernel	2.89	2.77	2.65	2.53	2.41	2.29	2.18	2.07	1.96	1.85
	GL	2.34	1.92	1.54	1.23	0.98	0.79	0.65	0.55	0.48	0.44
	Convolution	2.51	0.03	0.62	0.93	0.90	0.72	0.53	0.42	0.36	0.34
	$x_0 =$	0.01	0.12	0.23	0.34	0.45	0.56	0.67	0.78	0.89	1
$\Gamma(1.5,1)$	Kernel	0.45	0.35	0.42	0.34	0.24	0.15	0.10	0.08	0.08	0.09
	GL	0.18	0.02	0.02	0.02	0.07	0.16	0.02	0.06	0.04	0.17
	Convolution	0.74	0.30	0.08	0.05	0.01	0.05	0.05	0.04	0.04	0.05
	$x_0 =$	0.01	0.03	0.05	0.07	0.09	0.12	0.14	0.16	0.18	0.2
$\mathcal{U}([0,1])^2$	Kernel	17.56	6.06	2.43	0.61	0.36	0.82	0.98	q0.97	0.89	0.77
	GL	13.98	1.86	1.03	1.43	1.08	0.72	1.02	0.96	0.84	0.88
	Convolution	1.25	1.75	1.54	1.06	0.22	0.37	0.26	0.41	0.41	0.40
	$x_0 =$	0.01	0.12	0.23	0.34	0.45	0.56	0.67	0.78	0.89	1
$\lvert\mathcal{N}(0,1)\rvert$	Kernel	38.49	18.77	7.21	3.67	3.14	3.01	2.83	2.65	2.55	2.52
	GL	36.86	7.82	5.50	5.22	4.72	4.94	4.03	3.68	3.32	2.78
	Convolution	32.27	8.28	6.13	5.23	4.16	3.59	2.88	2.30	2.08	2.04
$\mathcal{G}P(4,2,0)$	Kernel	159.80	62.81	30.28	14.75	6.38	2.66	3.37	4.61	5.25	5.39
	GL	91.38	8.04	6.72	6.60	6.35	5.97	5.68	5.40	5.08	4.83
	Convolution	48.00	10.04	6.05	4.85	3.89	3.18	2.65	2.30	2.02	1.83

Table 1. $\sqrt{\text{MSE}(x_0)} \times 100$ for 10 values of x_0 near zero, 400 repetitions, four distributions and the three estimators, $n = 1000$.

multiplied by 100, where $\hat{f}_{(u)}$ denotes the estimator of f under study for the uth path generated. Clearly, the method performs very well near the origin, and the orders become the same when getting far from the problematic point.

We also give in Figure 3 plots of 25 estimated curves with the three methods for the $\Gamma(0.5,1)$ and the Pareto $\mathcal{G}P(4,1,0)$ densities. These variability bands show that the answers of the methods for samples of size 1000 are very stable.

5. Discussion and concluding remarks.

In this paper, we propose a new type of kernel estimators fitted to upper or lower bounded data, and in particular nonnegative data.

We prove that we can develop the complete theory of pointwise nonparametric density estimation in this context: bias-variance decomposition, higher order kernels for fitting functional regularities, adaptive bandwidth selection. Near zero, our kernel estimators are free of boundary effects.

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

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

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 inequality (unbounded random variables).

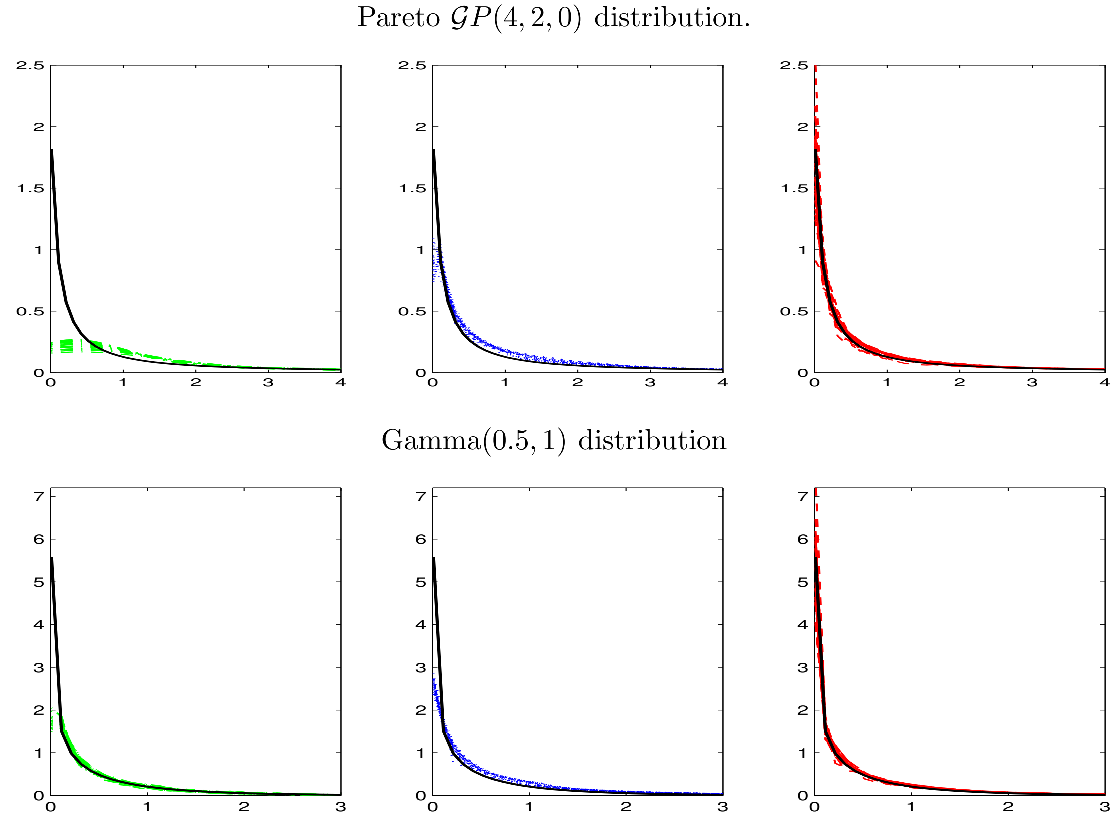

FIGURE 3. True density (bold black) and 25 estimated densities with the three methods: standard kernel (left), GL method (middle), our method (right), $n = 1000$.

In our numerical simulations, we have implemented the pointwise adaptive method. To compare our estimators to more standard kernel estimators, we have described how the adaptive method of Goldenschluger and Lespki (2011) 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 (7) and numerically more stable.

Lastly, let us mention that it is possible to study the integrated mean squared error of the estimator and propose global adaptive bandwidth selection (see Comte and Genon-Catalot (2011)).

6. PROOFS

6.1. **A useful Lemma.**

Lemma 6.1.
- *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,*
$$< K_m^{(1)}, K_m^{(2)} >= \int_0^{+\infty} K_m^{(1)}(u)K_m^{(2)}(u)du = \sqrt{m}(1/\sqrt{2\pi(v_1+v_2)})(1+o(1)).$$

Proof. 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:
$$D_m = \int_{\mathbb{R}} |K^*(s/\sqrt{m})|^{\alpha m} ds - \int_{\mathbb{R}} e^{-\alpha v s^2/2} ds \to 0 \tag{20}$$
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 (20) to study D_m. Now, we can follow closely the proof given in Feller (1971, p.516). We have $\varphi(t) = 1 - vt^2/2 + o(t^2)$. Using that $\varphi(s/\sqrt{m}) = 1 - vs^2/(2m) + o(1/m)$, 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:
$$\int_{\mathbb{R}} |v_m(s)| ds = \int_{|s|\leq a} \ldots + \int_{a<|s|<\delta\sqrt{m}} \ldots + \int_{|s|\geq\delta\sqrt{m}} \ldots \tag{21}$$
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 (21) 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 (20) 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

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

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

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

We develop K_m^2. By Lemma 6.1, as m tends to infinity, for $1 \le i, j \le L$,

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

The result follows. □

6.3. **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, $\gamma > 0, m \ge 2$,

$$\mu_\ell^{(j)} = \int_0^{+\infty} (u-1)^\ell K^{(j)}(u)du, \quad \mu_{\ell,m}^{(j)} = \int_0^{+\infty} (u-1)^\ell K_m^{(j)}(u)du, \tag{22}$$

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

$$\nu_\gamma^{(j)} = \int_0^{+\infty} |u-1|^\gamma K^{(j)}(u)du, \quad \nu_{\gamma,m}^{(j)} = \int_0^{+\infty} |u-1|^\gamma K_m^{(j)}(u)du. \tag{23}$$

Lemma 6.2. *Assume that the density $K^{(j)}$ satisfies* (H1) *and* (H3)*. Then, for all $\alpha \le \gamma$,*

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

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 $\mu_k^{(j)}, k \le \ell$. In particular, if $\mu_{2\ell}^{(j)} < +\infty$ for some integer ℓ, then, $\mu_{2\ell,m}^{(j)} \le m^{-\ell}C(\mu_{2k}^{(j)}, k \le \ell)$.

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 6.2 (see Tsybakov (2009)).□

Proof of Lemma 6.2. 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 $4/\alpha > 1$, $\nu_{\alpha,m} = \mathbb{E}|\bar{U}_m - 1|^\alpha \le \left(\mathbb{E}(\bar{U}_m - 1)^4\right)^{\alpha/4}$. 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 $\alpha/4 < 1/2 < 1$, $\left(\mathbb{E}(\bar{U}_m - 1)^4\right)^{\alpha/4} \leq (6v^2)^{\alpha/4} m^{-\alpha/2} + \mu_4^{\alpha/4} m^{-3\alpha/4}$. 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 $\alpha - 1 \geq \alpha/2$. □

6.4. **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: $i\kappa_1 = \varphi'(0) = \psi'(0) = 0$, $-\kappa_2 = \varphi''(0) = \psi''(0) = -v$, 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

$$\psi_a'(0) = 0 = \varphi_a'(0). \tag{24}$$

For $j \geq 2$,

$$\psi_a^{(j)}(0) = i^j \kappa_{j,a} = i^j \int_0^{+\infty} x^j n_a(x)dx = i^j \frac{1}{a^{j-1}}\kappa_j. \tag{25}$$

Cumulants and (centered) moments are linked by the induction formula (see *e.g.* Abramowitz and Stegun(1964)) (using that $\kappa_1 = 0$, see (25)):

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

We deduce, by a tedious but elementary induction: $\mu_{n,a} = \sum_{\ell=k(n)}^{n-1} c(\ell)/a^\ell$ 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. □

6.5. **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 $k = 1, \dots, \ell - 1$, $\sum_{j=1}^{\ell} \binom{\ell}{j}(-1)^{j+1} j^k = 0$, which, in turn, holds, as, for $k = 1, \dots, \ell - 1$,

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

6.6. **Proof of Theorem 3.1.**

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

$$B_m f(x) = \mathbb{E}\hat{f}_m(x) - f(x), \quad B_{m,m'} f(x) = \mathbb{E}\hat{f}_{m,m'}(x) - f(x). \tag{26}$$

The following relation between bias terms holds.

Lemma 6.3. $B_{m,m'} f(x) = B_{m'} f(x) + \int_0^{+\infty} K_{m'}(u) B_m f(xu) du.$

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.$\square$

Lemma 6.4. *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,*

$$||K_m \odot K_{m'}||_\infty \le \sqrt{m \wedge m'} (\sum_{i=1}^{L} |\alpha_i|)(\sum_{i=1}^{L} \frac{|\alpha_i|}{\sqrt{2\pi v_i}})(1 + o(1)). \tag{27}$$

Proof. Consider $U_k^{(i)}, k=1,\dots,m, V_\ell^{(j)}, \ell=1,\dots,m'$ 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 $\bar{U}_m^{(i)} = (U_1^{(i)}+\dots+U_m^{(i)})/m$ and $I_m^{(i)} = \mathbb{E}(1/\bar{U}_m^{(i)}) = 1+O(1/m)$. Analogously, $K_{m'}^{(j)}$ is the density of $\bar{V}_{m'}^{(j)} = (V_1^{(j)}+\dots+V_{m'}^{(j)})/m'$ 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 6.1, 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 6.5. *We have:*

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

where the constant $C(\alpha,v)$ (see (17)) does not depend on the density f.

Proof. We have:

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

we conclude using Lemma 6.4 and the fact that $K_m^{(i)} \odot K_{m'}^{(j)}(u)$ are densities. □

6.6.2. *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 6.3, for all $m,m'\in\mathcal{M}_n$,

$$|f_{m'}(x_0)-f_{m,m'}(x_0)| = |\int_0^{+\infty} B_m f(x_0u)K_{m'}(u)du|.$$

Therefore, using that each $K_{m'}^{(i)}$ is a density, Lemma 6.2 and Proposition 3.2, we obtain:

$$\begin{aligned}
|f_{m'}(x_0)-f_{m,m'}(x_0)| &\leq \sum_{i=1}^{L}|\alpha_i|\int_0^{+\infty}|(B_mf)(xu)|K_{m'}^{(i)}(u)du\\
&\leq \frac{SC_0}{\lfloor\beta_0\rfloor!}\frac{x_0^{\beta_0}}{m^{\beta_0/2}}\sum_{i=1}^{L}|\alpha_i|\int u^{\beta}K_{m'}^{(i)}(u)du\\
&\leq \frac{SC_0 2^{\beta_0}}{\lfloor\beta_0\rfloor!}\frac{x_0^{\beta_0}}{m^{\beta_0/2}}\sum_{i=1}^{L}|\alpha_i|(\int K_{m'}^{(i)}(u)du+\nu_{\beta_0,m'}^{(i)})\\
&\leq \frac{SC_0 2^{\beta_0}}{\lfloor\beta_0\rfloor!}\frac{x_0^{\beta_0}}{m^{\beta_0/2}}\sum_{i=1}^{L}|\alpha_i|(1+c_i(\beta_0)(m')^{-\beta_0/2})\leq C'(\beta_0,\alpha)\frac{x_0^{\beta_0}+1}{m^{\beta_0/2}},
\end{aligned}$$

where

$$C'(\beta_0,\alpha)=\frac{SC_0 2^{\beta_0}}{\lfloor\beta_0\rfloor!}\sum_{i=1}^{L}|\alpha_i|(1+c_i(\beta_0)),$$

and $C_0=C_0(\beta_0)$, see (8). 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:

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

Now, we can prove the following Lemmas:

Lemma 6.6. *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 6.7. *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. □

6.6.3. *Proof of Lemma 6.6.* First we write,

(28)
$$\mathbb{E}\left(\sup_{m'}\left((\hat{f}_{m'}(x_0)-f_{m'}(x_0))^2-\frac{V(m',x_0)}{6}\right)_+\right)\leq\sum_{m\in\mathcal{M}_n}\mathbb{E}\left(\left((\hat{f}_m(x_0)-f_m(x_0))^2-\frac{V(m,x_0)}{6}\right)_+\right).$$

Then

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

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

Lemma 6.8. *Let $T_1,\dots,T_n$ be independent random variables and $S_n(T)=\sum_{i=1}^n[T_i-\mathbb{E}(T_i)]$. Then, for $\eta>0$,*

$$\begin{aligned} P(|S_n(T)-\mathbb{E}(S_n(T))|\ge n\eta) &\le 2\exp\left(\frac{-n\eta^2/2}{v^2+b\eta}\right)\\ &\le 2\min\left(\exp\left(-\frac{n\eta^2}{4v^2}\right),\exp\left(-\frac{n\eta}{4b}\right)\right), \end{aligned}\tag{29}$$

where

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

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 $|K_m|\le\sum_{i=1}^L|\alpha_i|K_m^{(i)}$ where, for all u,

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

by Lemma 6.1. Hence,

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

Using the variance bound, we choose, for $C(\alpha,v)$ given by (17)

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

We find,

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

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

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

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

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

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. □

6.6.4. *Proof of Lemma 6.7.* The proof follows the same line. The new elements are to be taken from (27) in Lemma 6.4 and from Lemma 6.5.

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.

Cheng, M.Y. (1997). Boundary-aware estimators of integrated density derivative. *Journal of the Royal Statistical Society Ser. B*, **59**, 191-203.

Cheng, M.Y., Fan, J. and Marron, J.S. (1997). On automatic boundary corrections. *The Annals of Statistics*, **25**, 1691-1708.

Cline, D.B.H. and Hart, J.D. (1991). Kernel estimation of densities of discontinuous derivatives. *Statistics*, **22**, 69-84.

Comte, F. and Genon-Catalot, V. (2011). Density estimation for nonnegative random variables. *Preprint* MAP5 2011-06.

Cowling, A. and Hall, P. (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **58**, no. 3, 551563.

Devroye, L. (1989). The double kernel method in density estimation. *Ann. Inst. H. Poincaré 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.

Efromovich, S. (1999) *Nonparametric curve estimation. Methods, theory, and applications.* Springer Series in Statistics. Springer-Verlag, New York.

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

Gasser, T., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society, Ser. B*, **47**, 238-252.

Goldenshluger A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, no. 3, 1608-1632.

Hall, P. (1983). On near neighbour estimates of a multivariate density. *J. Multivariate Anal.* **13**, no. 1, 2439.

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.

Hall, P. and Park, B.U. (2002) New methods for bias correction at endpoints and boundaries. *Ann. Statist.* **30**, no. 5, 14601479.

Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Statistics and computing*, **3**, 135-146.

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

Karunamuni, R. J. and Alberts, T. (2005). On boundary correction in kernel density estimation. *Stat. Methodol.* **2**, no. 3, 191212.

Karunamuni, R. J. and Alberts, T. (2006). A locally adaptive transformation method of boundary correction in kernel density estimation. *J. Statist. Plann. Inference* **136**, no. 9, 29362960.

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.

Killmann, F. and von Collani, E. (2001). A note on the convolution of the uniform and related distributions and their use in quality control. *Economic Quality Control* **16**, no. 1, 17-41.

Mack, Y. P. and Rosenblatt, M. (1979). Multivariate k-nearest neighbor density estimates.*J. Multivariate Anal.* **9**, no. 1, 115.

Marron, J.S. and Ruppert, D. (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society Ser. B*, **56**, 653-671.

Rényi, A. (1970). *Probability Theory.* North-Holland, Amsterdam.

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

Silverman, B. W. (1986). *Density estimation for statistics and data analysis.* Monographs on Statistics and Applied Probability. Chapman and Hall, London.

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

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.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, London, Chapman and Hall.

Wand, M.P., Marron, J.S. and Ruppert, D. (1991). Transformations in density estimation (with discussion). *Journal of the American Statistical Association*, **86**, 343-361.

Zhang, S. and Karunamuni, R.J. (1998). On kernel density estimation near endpoints. *J. Statist. Plann. Inference* **70**, no. 2, 301-316.

Zhang , S. and Karunamuni, R.J. (2000). On nonparametric density estimation at the boundary. *J. Nonparametr. Statist.* **12**, no. 2, 197-221

Zhang, S., Karunamuni, R.J. and Jones, M.C. (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association*, **448**, 1231-1241.