# ADAPTIVE KERNEL ESTIMATION OF THE LÉVY DENSITY

MÉLINA BEC*, CLAIRE LACOUR**

ABSTRACT. This paper is concerned with adaptive kernel estimation of the Lévy density $N(x)$ for pure jump Lévy processes. The sample path is observed at $n$ discrete instants in the "high frequency" context ($\Delta = \Delta_n$ tends to zero while $n\Delta_n$ tends to infinity). We construct a collection of kernel estimators of the function $g(x) = xN(x)$ and propose a method of local adaptive selection of the bandwidth. We provide an oracle inequality and a rate of convergence for the quadratic pointwise risk. This rate is proved to be the optimal minimax rate. We give examples and simulation results for processes fitting in our framework.

KEYWORDS. Adaptive Estimation; High frequency; Pure jump Lévy process; Nonparametric Kernel Estimator.

May 21, 2012

## 1. INTRODUCTION

Consider $(L_t, t \geq 0)$ a real-valued Lévy process with characteristic function given by:

$$\psi_t(u) = \mathbb{E}(\exp iuL_t) = \exp\left(t \int_{\mathbb{R}} (e^{iux} - 1)N(dx)\right). \tag{1}$$

We assume that the Lévy measure admits a density $N$ and that the function $g(x) = xN(x)$ is integrable. Under these assumptions, $(L_t, t \geq 0)$ is a pure-jump Lévy process without drift and with finite variation on compact sets. Moreover $\mathbb{E}(|L_t|) < \infty$ (see Bertoin (1996)). Suppose that we have discrete observations $(L_{k\Delta}, k = 1, ..., n)$ with sampling interval $\Delta$. Our aim in this paper is the nonparametric adaptive kernel estimation of the function $g(x) = xN(x)$ based on these observations under the asymptotic framework $n$ tends to $\infty$. This subject has been recently investigated by several authors. Figueroa-López and Houdré (2006) use a penalized projection method to estimate the Lévy density on a compact set separated from 0. Other authors develop an estimation procedure based on empirical estimations of the characteristic function $\psi_\Delta(u)$ of the increments $(Z_k^\Delta = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \ldots, n)$ and its derivatives followed by a Fourier inversion to recover the Lévy density. For low frequency data ($\Delta$ is fixed), we can quote Watteel and Kulperger (2003), or Jongbloed and van der Meulen (2006) for a parametric study. Still in the low frequency framework, Neumann and Reiß (2009) estimate $\nu(x) = x^2N(dx)$ in the more general case with drift and volatility, and Comte and Genon-Catalot (2010) use model selection to build an adaptative estimator. Belomestny (2011) addresses the issue of inference for time-changed Lévy processes with results in term of uniform and pointwise distance.

* UMR CNRS 8145 MAP5, Université Paris Descartes, ** Laboratoire de mathématiques, Université Paris-Sud 11.

In the high frequency context, which is our concern in this paper, the problem is simpler since, for any fixed $u$, $\psi_\Delta(u) \to 1$ when $\Delta \to 0$. This implies that $\psi_\Delta(u)$ need not to be estimated and can simply be replaced by 1 in the estimation procedures. This is what is done in Comte and Genon-Catalot (2009). These authors start from the equality:

$$\mathbb{E}\left[Z_k^\Delta e^{iuZ_k^\Delta}\right] = -i\psi'_\Delta(u) = \Delta\psi_\Delta(u)g^*(u), \tag{2}$$

obtained by differentiating (1). Here $g^*(u) = \int e^{iux}g(x)dx$ is the Fourier transform of $g$, well defined since we assume $g$ integrable. Then, as $\psi_\Delta(u) \simeq 1$, equation (2) writes $\mathbb{E}\left[Z_k^\Delta e^{iuZ_k^\Delta}\right] \simeq \Delta g^*(u)$. This gives an estimator of $g^*(u)$ as follows:

$$\frac{1}{n\Delta}\sum_{k=1}^n Z_k^\Delta e^{iuZ_k^\Delta}.$$

Now, to recover $g$ the authors apply Fourier inversion with cutoff parameter $m$. Here, we rather introduce a kernel to make inversion possible:

$$\frac{1}{n\Delta}\sum_{k=1}^n Z_k^\Delta K^*(uh)e^{iuZ_k^\Delta} \tag{3}$$

which is in fact the Fourier transform of $1/(nh\Delta)\sum_{k=1}^n Z_k^\Delta K((x - Z_k^\Delta)/h)$. At the end, in the high frequency context, a direct method without Fourier inversion can be applied. Indeed, a consequence of (2) is that the empirical distribution:

$$\hat\mu_n(dz) = \frac{1}{n\Delta}\sum_{k=1}^n Z_k^\Delta \delta_{Z_k^\Delta}(dz)$$

weakly converges to $g(z)dz$. This suggests to consider kernel estimators of $g$ of the form

$$\hat g_h(x) = K_h \star \hat\mu_n(x) = \frac{1}{n\Delta}\sum_{k=1}^n Z_k^\Delta K_h(x - Z_k^\Delta) \tag{4}$$

where $K_h(x) = (1/h)K(x/h)$ and $K$ is a kernel such that $\int K = 1$. Below, we study the quadratic pointwise risk of the estimators $\hat g_h(x)$ and evaluate the rate of convergence of this risk as $n$ tends to infinity, $\Delta = \Delta_n$ tends to 0 and $h = h_n$ tends to 0. This is done under Hölder regularity assumptions for the function $g$. Note that a pointwise study involving a kernel estimator can be found in van Es et al. (2007) for more specific compound Poisson processes, but the estimator is different from ours, as well as the observation scheme. In Figueroa-López (2011) a pointwise central limit theorem is given for the estimation of the Lévy density, as well as confidence intervals. Still in the high frequency context, we can cite Duval (2012) for the estimation of a compound Poisson process with low conditions on $\Delta$, but for integrated distance.

In this paper, we study local adaptive bandwidth selection (which the previous authors do not consider). For a given non-zero real $x_0$, we select a bandwidth $\hat h(x_0)$ such that the resulting adaptive estimator $\hat g_{\hat h(x_0)}(x_0)$ automatically reaches the optimal rate of convergence corresponding to the unknown regularity of the function $g$. The method of bandwidth selection follows the scheme developped by Goldenshluger and Lepski (2011) for density estimation. Moreover our contribution includes a lower bound result to prove the optimality of the rate for this pointwise estimation.

In Section 2, we give notations and assumptions. In Section 3, we study the pointwise mean square error (MSE) of $\hat{g}_h(x_0)$ given in (4) for $g$ belonging to a Hölder class of regularity $\beta$ and we present the bandwidth selection method together with both lower and upper risk bound for our adaptive estimator. The rate of convergence of the risk is $(\log(n\Delta)/n\Delta)^{2\beta/2\beta+1}$ which is expected in adaptive pointwise context. Examples and simulations in our framework are discussed in Section 4 and proofs are gathered in Section 5.

## 2. Notations and assumptions

We present the assumptions on the kernel $K$ and on the function $g$ required to study the estimator given by (4). First, we set some notations. For any functions $u, v$, we denote by $u^*$ the Fourier transform of $u$, $u^*(y) = \int e^{iyx}u(x)dx$ and by $\|u\|$, $< u, v >$, $u \star v$ the quantities

$$\|u\|^2 = \int |u(x)|^2 dx,$$

$$< u, v >= \int u(x)\overline{v}(x)dx \text{ with } z\overline{z} = |z|^2 \text{ and } u \star v(x) = \int u(y)v(x-y)dy.$$

For a positive real $\beta$, $\lfloor \beta \rfloor$ denotes the largest integer strictly smaller than $\beta$. Let us also define the following functional space:

**Definition 2.1.** *(Hölder class) Let $\beta > 0$, $L > 0$ and let $l = \lfloor \beta \rfloor$. The Hölder class $\mathcal{H}(\beta, L)$ on $\mathbb{R}$ is the set of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that derivative $f^{(l)}$ exists and verifies:*

$$|f^{(l)}(x) - f^{(l)}(y)| \leq L|x-y|^{\beta-l}, \quad \forall x, y \in \mathbb{R}.$$

We can now define the assumptions concerning the target function $g$:

**G1:** $g \in \mathbb{L}^2$
**G2:** $g^*$ is differentiable almost everywhere and its derivative belongs to $\mathbb{L}^1$
**G3($p$):** For $p$ integer, $\int |x|^{p-1}|g(x)|dx < \infty$
**G4($\beta$):** $g \in \mathcal{H}(\beta, L)$
**G5:** $g'$ exists and is uniformly bounded

The first assumption is natural to use Fourier analysis, as well as G3(1). Assumption G3($p$) ensures that $\mathbb{E}|Z_1|^p < \infty$. G4 is a classical regularity assumption in nonparametric estimation; it allows to quantify the bias (see Tsybakov (2009)). Note that G5 implies that $f \in \mathcal{H}(1, L')$ so we can assume $\beta \geq 1$.

Now let us describe which kind of kernel we choose for our estimator. For $m \geq 1$ an integer, we say that $K : \mathbb{R} \to \mathbb{R}$ is a kernel of order $m$ if functions $u \mapsto u^j K(u), j = 0, 1, ..., m$ are integrable and satisfy

$$\int K(u)du = 1, \qquad \int u^j K(u)du = 0, \quad j \in \{1, ..., m\}. \tag{5}$$

Let us define the following conditions

**K1:** $K$ belongs to $\mathbb{L}^1 \cap \mathbb{L}^2 \cap \mathbb{L}^\infty$ and $K^* \in \mathbb{L}^1$
**K2($\beta$):** The kernel $K$ is of order $l = \lfloor \beta \rfloor$ and $\int |x|^\beta |K(x)|dx < +\infty$

These assumptions are standard when working on problems of estimation by kernel methods. Note that there is a way to build a kernel of order $l$. Indeed, let $u$ be a bounded integrable function such that $u \in \mathbb{L}^2$, $u^* \in \mathbb{L}^1$ and $\int u(y)dy = 1$, and set for any given integer $l$,

$$K(t) = \sum_{k=1}^{l} \binom{l}{k} (-1)^{k+1} \frac{1}{k} u\left(\frac{t}{k}\right) \tag{6}$$

The kernel $K$ defined by (6) is a kernel of order $l$ which also satisfies K1 (see Kerkyacharian et al. (2001) and Goldenshluger and Lepski (2011)). As usual, we define $K_h$ by

$$\forall x \in \mathbb{R} \qquad K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right).$$

In all the following we fix $x_0 \in \mathbb{R}$, $x_0 \neq 0$.

## 3. Risk bound

### 3.1. Risk bound for a fixed bandwidth.

In this subsection, the bandwidth $h$ is fixed, thus we omit the subscript $h$ for the sake of simplicity: we denote $\hat{g} = \hat{g}_h$. The usual bias variance decomposition of the MSE yields:

$$MSE(x_0, h) := \mathbb{E}[(\widehat{g}(x_0) - g(x_0))^2] = \mathbb{E}[(\widehat{g}(x_0) - \mathbb{E}[\widehat{g}(x_0)])^2] + (\mathbb{E}[\widehat{g}(x_0)] - g(x_0))^2$$

But the bias needs further decomposition:

$$b(x_0)^2 := (\mathbb{E}[\widehat{g}(x_0)] - g(x_0))^2 \quad \leq \quad 2b_1(x_0)^2 + 2b_2(x_0)^2$$

with the usual bias,

$$b_1(x_0) = K_h \star g(x_0) - g(x_0),$$

and the bias resulting from the approximation of $\psi_\Delta(u)$ by 1,

$$b_2(x_0) = \mathbb{E}[\widehat{g}(x_0)] - K_h \star g(x_0).$$

We can provide the following bias bound:

**Lemma 3.1.** *Under G3(1), G4($\beta$), G5 and if the kernel $K$ satisfies K1 and K2($\alpha$) with $\alpha \geq \beta$*

$$|b(x_0)|^2 \quad \leq \quad c_1 h^{2\beta} + c'_1 \Delta^2$$

*with $c_1 = 2\left(L/\lfloor\beta\rfloor! \int |K(v)||v|^\beta dv\right)^2$ and $c'_1 = 2(2\|g'\|_\infty \|g\|_1 \|K\|_1)^2$.*

Moreover, the variance is controlled as follows:

**Lemma 3.2.** *Under G1 and G2, and if the kernel satisfies K1, we have*

$$\mathrm{Var}[\widehat{g}(x_0)] \leq \frac{1}{nh\Delta} \frac{\|K\|_2^2}{2\pi} (\|(g^*)'\|_1 + \|g^*\|_2^2 \Delta) \leq c_2 \frac{1}{nh\Delta} + c'_2 \frac{1}{nh}$$

*with $c_2 = \|(g^*)'\|_1 \|K\|_2^2/(2\pi)$ and $c'_2 = \|K\|_2^2 \|g\|_2^2$.*

Lemmas 3.1 et 3.2 lead us to the following risk bound:

**Proposition 3.1.** *Under G1, G2, G3(1), G4($\beta$), G5 and if $K$ satifies K1 and K2($\alpha$) with $\alpha \geq \beta$, we have*

$$MSE(x_0, h) \leq c_1 h^{2\beta} + c_2 \frac{1}{nh\Delta} + c'_2 \frac{1}{nh} + c'_1 \Delta^2. \tag{7}$$

Recall that $\Delta = \Delta_n$ is such that $\lim_{n\to+\infty} \Delta_n = 0$, thus $1/nh$ is negligible compared to $1/nh\Delta$. For the two first terms the optimal choice of $h$ is $h_{opt} \propto ((n\Delta)^{-\frac{1}{2\beta+1}})$ and the associated rate has order $O\left((n\Delta)^{-\frac{2\beta}{2\beta+1}}\right)$. Next, a sufficient condition for $\Delta^2 \leq (n\Delta)^{-\frac{2\beta}{2\beta+1}}$ for all $\beta$ is

$$\Delta = O(n^{-1/3}). \tag{8}$$

**Proposition 3.2.** *Under the assumptions of Proposition 3.1 and under condition (8), the choice $h_{opt} \propto ((n\Delta)^{-\frac{1}{2\beta+1}})$ minimizes the risk bound (7) and gives $MSE(x_0, h_{opt}) = O((n\Delta)^{-\frac{2\beta}{2\beta+1}})$. As a consequence $\mathbb{E}[(\widehat{g}(x_0)/x_0 - N(x_0))^2] = O((n\Delta)^{-\frac{2\beta}{2\beta+1}})$.*

We can link this result to the one of Figueroa-López (2011) who proves that his projection estimator $\widehat{N}$ is such that $(\widehat{N}(x_0) - N(x_0))(n\Delta)^{\alpha}$ tends to a normal distribution for any $0 < \alpha < \beta/(2\beta + 1)$.

3.2. **Lower bound.** The rate obtained in Proposition 3.2 turns out to be the optimal minimax rate of convergence over the class $\mathcal{H}(\beta, L)$. Indeed we prove the following lower bound:

**Theorem 3.1.** *Assume $\Delta = O(1)$ and $\Delta^{-1} = O(n)$. Let $x_0 \neq 0$. There exists $C > 0$ such that for any estimator $\hat{g}_n(x_0)$ based on observations $Z_1^{\Delta}, \ldots, Z_n^{\Delta}$, and for $n$ large enough,*

$$\sup_{g\in\mathcal{H}(\beta,L)} \mathbb{E}_g \left[(\hat{g}_n(x_0) - g(x_0))^2\right] \geq C(n\Delta)^{-\frac{2\beta}{2\beta+1}}.$$

Obviously, the result is also true replacing $g$ by the Lévy density $N$.

3.3. **Bandwidth selection.** As $\beta$ is unknown, we need a data-driven selection of the bandwidth. We follow ideas given in Goldenshluger and Lepski (2011) for density estimation. We introduce a set of bandwidth of the form $H = \{\frac{j}{M}, 1 \leq j \leq M\}$ with $M$ an integer to be specified later. Actually it is sufficient to control $\sum_{h\in H} h^{-w}$ for some $w$ so that more general set of bandwiths are possible. We set:

$$V(h) = C_0 \frac{\log(n\Delta)}{nh\Delta}$$

with $C_0$ to be specified later. Note that $V(h)$ has the same order as the variance multiplied by $\log(n\Delta)$. We also define $\hat{g}_{h,h'}(x_0) = K_{h'} \star \hat{g}_h(x_0) = K_h \star \hat{g}_{h'}(x_0)$. This auxiliary estimator can also be written

$$\hat{g}_{h,h'}(x_0) = \frac{1}{n\Delta} \sum_{k=1}^{n} Z_k^{\Delta} K_{h'} \star K_h(x_0 - Z_k^{\Delta}).$$

Lastly we set

$$A(h, x_0) = \sup_{h'\in H} \left[|\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0)|^2 - V(h')\right]_+.$$

The adaptive bandwidth $h$ is chosen as follows:

$$\hat{h} = \hat{h}(x_0) \in \arg\min_{h\in H}\{A(h, x_0) + V(h)\}.$$

We can state the following oracle inequality.

**Theorem 3.2.** *We use a kernel satisfying* $K1$ *and* $M = O((n\Delta)^{1/3})$. *Assume that* $g$ *satisfies G1, G2, G3(5) and take*

$$C_0 = \frac{c}{2\pi}\|K\|^2 \left(\|(g^*)'\|_1 + \Delta\|g^*\|_2^2\right) \tag{9}$$

*with* $c \geq 16\max(1, \|K\|_\infty)$. *Then*

$$\mathbb{E}[|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2] \leq C\left\{\inf_{h\in H}\left\{\|g - \mathbb{E}[\hat{g}_h]\|_\infty^2 + V(h)\right\} + \frac{\log(n\Delta)}{n\Delta}\right\}$$

Thus our estimator $\hat{g}_{\hat{h}}$ has a risk as good as any of the collection $(\hat{g}_h)_{h\in H}$, up to a logarithmic term. The definition of the estimator uses $\|(g^*)'\|_1$ and $\|g^*\|_2^2$, but these quantities can be estimated with a preliminar estimator of $g^*$, as given in (3) (see Birgé and Massart (1997) for an example of such a plug-in in a penalty term).

Let us now conclude with the consequence of this theorem in term of rate of convergence. As already explained, as we need assumption G5 to control the bias, we can assume $\beta \geq 1$. Then $h_{opt} \propto (\log(n\Delta)/n\Delta)^{1/(2\beta+1)} \geq (n\Delta)^{-1/3}$ belongs to $H$ as soon as $M$ is larger than a constant times $(n\Delta)^{1/3}$. Hence we can state the following corollary.

**Corollaire 3.1.** *Assume that* $g$ *satisfies G1, G2, G3(5), G4(*$\beta$*) with* $\beta \geq 1$ *and G5. We choose a kernel satisfying K1 and K2(*$\alpha$*) with* $\alpha \geq \beta$, *and* $M = \lfloor (n\Delta)^{1/3}\rfloor$. *Take* $C_0$ *as in Theorem 3.2. Then, if* $n^{-1} \ll \Delta \leq Cn^{-1/3}$,

$$\mathbb{E}[|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2] = O\left((\log(n\Delta)/n\Delta)^{-\frac{2\beta}{2\beta+1}}\right)$$

Then the price to pay to adaptivity is a logarithmic loss in the rate. Nevertheless this phenomenon is known to be unavoidable in pointwise estimation (see Butucea (2001)). Thus $\hat{g}_{\hat{h}}(x_0)$ (resp. $\hat{g}_{\hat{h}}(x_0)/x_0$) is an adaptive estimator for $g(x_0)$ (resp. $N(x_0)$).

## 4. Examples and Simulations

We have implemented the estimation method for three different processes (listed in Examples 1-3 below) with the kernel described in (6) (with $l = 2$ and $u$ the Gaussian density). For the implementation, a difficulty is the proper calibration of the constant $c$ in (9). This is usually done by a large number of previous simulations. We have chosen $c = 1$ as the adequate value for a variety of models, kernels and number of observations. For clarity, we have computed the Mean Integrated Square Error (MISE) of the estimators. Figure 1 plots ten estimated curves corresponding to our three examples with in the first column $n = 10000$, $\Delta = 0.1$, and in the second $n = 100000$, $\Delta = 0.1$.

**Example 1.** Let $L_t = \sum_{i=1}^{N_t} Y_i$, where $(N_t)$ is a Poisson process with constant intensity $\lambda$ and $(Y_i)$ is a sequence of i.i.d random variables with density $f$ independent of the process $(N_t)$. Then, $(L_t)$ is a Lévy process with characteristic function

$$\psi_t(u) = \exp\left(\lambda t\int_{\mathbb{R}}(e^{iux} - 1)f(x)dx\right). \tag{10}$$

Its Lévy density is $N(x) = \lambda f(x)$ and thus $g(x) = \lambda x f(x)$. For our first example, we choose $\lambda = 1$ and $f$ such that $g(x) = xf(x) = (1/2)\sqrt{x/2}$ for $0 < x \le 2$. Then assumption G4(1/2) holds (on $(0, 2)$), but not G4($\beta$) for other $\beta$. Since $\beta$ is small, the rate of convergence is slow. The discontinuity in 2 damages the estimation as it can be seen in Figure 1.

**Example 2.** Let $\alpha > 0$, $\gamma > 0$. The Lévy-Gamma process $(L_t)$ with parameters $(\gamma, \alpha)$ is such that, for all $t > 0$, $L_t$ has Gamma distribution with parameters $(\gamma t, \alpha)$, i.e the density:

$$\frac{\alpha^{\gamma t}}{\Gamma(\gamma t)} x^{\gamma t - 1} e^{-\alpha x} \mathbb{1}_{x \ge 0}.$$

The Lévy density is $N(x) = \gamma x^{-1} e^{-\alpha x} \mathbb{1}_{x>0}$ so that $g(x) = \gamma e^{-\alpha x} \mathbb{1}_{x>0}$ satisfies assumptions G1, G2 and G3($p$). Here we choose $\alpha = \gamma = 1$. This example allows to study the role of the discontinuity in 0, which invalidates assumptions G4-G5. We can observe that the estimation become very good if we move away from 0.

**Example 3.** For our third example, we also choose a compound Poisson process, but with $f$ the standard Gaussian density. Thus $g(x) = xf(x) = xe^{-x^2/2}/\sqrt{2\pi}$ and $g^*(u) = iue^{-u^2/2}$. Assumptions G1, G2, G3($p$),G5 hold for $g$. Moreover $g$ belongs to a Hölder class of regularity $\beta$ for all $\beta > 0$. Thus the rate is close to $(n\Delta/\log(n\Delta))^{-1}$, and the good performance of our estimator is visible on the figure.

## 5. Proofs

Let us first state two useful propositions (see Proposition 2.1 in Comte and Genon-Catalot (2010) and Proposition 2.1 in Comte and Genon-Catalot (2009) for a proof).

**Proposition 5.1.** *Denote by $P_\Delta$ the distribution of $Z_1^\Delta$ and define $\mu_\Delta(dx) = \Delta^{-1} x P_\Delta(dx)$. If $\int_{\mathbb{R}} |x| N(x) < \infty$, the distribution $\mu_\Delta$ has a density $h_\Delta$ given by*

$$h_\Delta(x) = \int g(x-y) P_\Delta(dy) = \mathbb{E} g(x - Z_1^\Delta).$$

**Proposition 5.2.** *Let $p \ge 1$ an integer such that $\int_{\mathbb{R}} |x|^{p-1} |g(x)| dx < \infty$. Then $\mathbb{E}(|Z_1^\Delta|^p) < \infty$ and $\mathbb{E}[(Z_1^\Delta)^p] = \Delta \int_{\mathbb{R}} x^{p-1} g(x) dx + o(\Delta)$. Moreover, if $g$ is integrable, $\mathbb{E}(|Z_1^\Delta|) \le 2\Delta \|g\|_1$.*

### 5.1. Proof of Lemma 3.1.

First, we study $b_2(x_0)$ using Proposition 5.1:

$$\begin{aligned} b_2(x_0) &= \frac{1}{h\Delta} \mathbb{E}\left[ Z_1^\Delta K\left( \frac{x_0 - Z_1^\Delta}{h} \right) \right] - \frac{1}{h} \int K\left( \frac{x_0 - u}{h} \right) g(u) du \\ &= \frac{1}{h} \int K\left( \frac{x_0 - u}{h} \right) \mathbb{E}[g(u - Z_1^\Delta)] du - \frac{1}{h} \int K\left( \frac{x_0 - u}{h} \right) g(u) du \\ &= \frac{1}{h} \int K\left( \frac{x_0 - u}{h} \right) \mathbb{E}[g(u - Z_1^\Delta) - g(u)] du. \end{aligned}$$

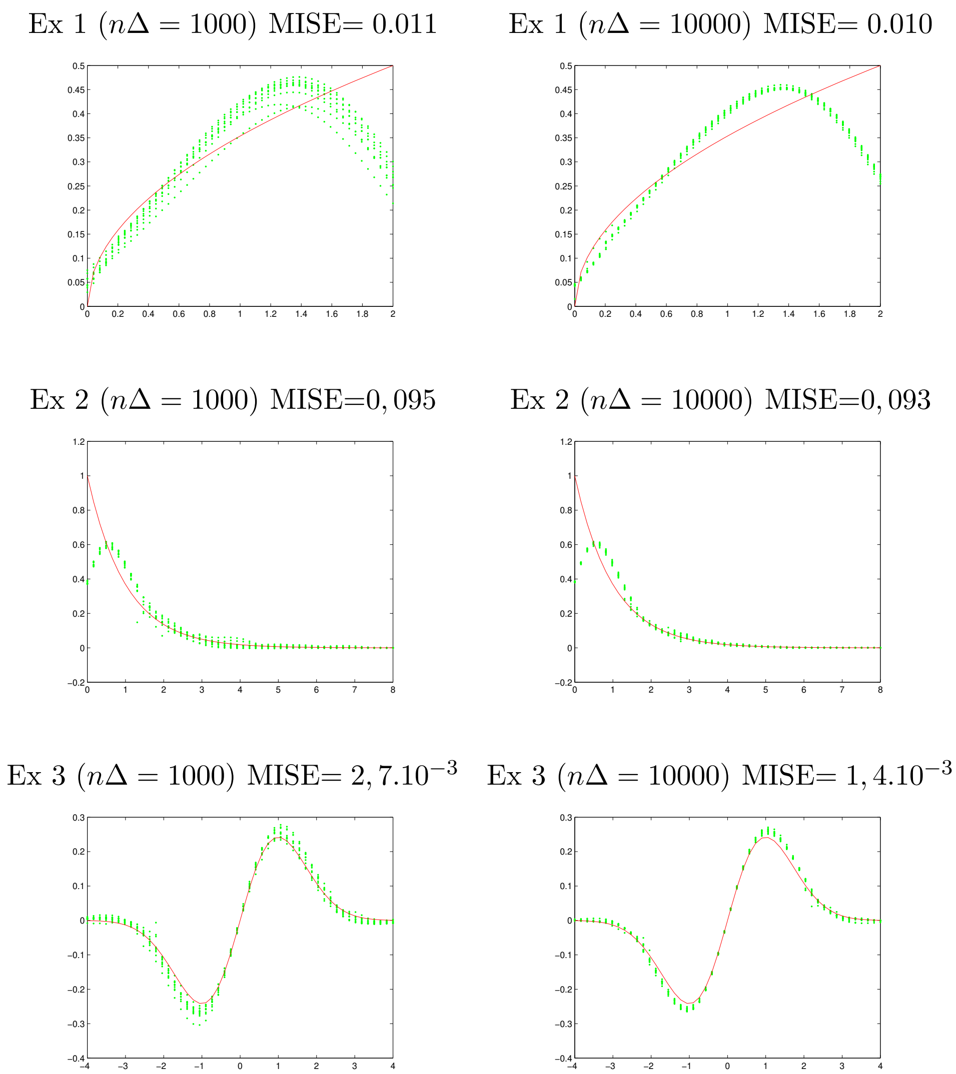


FIGURE 1. Function $g$ (solid line) and estimators $\hat{g}_{\hat{h}}$ (dotted lines).

Now, applying the mean value theorem to $g$, we get

$$\begin{aligned} |b_2(x_0)| &= \left| \frac{1}{h} \int K\left(\frac{x_0 - u}{h}\right) \mathbb{E}[-Z_1^\Delta g'(u_{Z_1})] du \right| \text{ with } u_{Z_1} \in [u - Z_1, u\ ] \\ &\leq \|g'\|_\infty \|K\|_1 \mathbb{E}\left|Z_1^\Delta\right|. \end{aligned}$$

From the results of Proposition 5.2 we obtain

$$|b_2(x_0)| \leq \ 2\|g'\|_\infty \|K\|_1 \|g\|_1 \Delta. \tag{11}$$

Now we study $b_1(x_0) = K_h \star g(x_0) - g(x_0)$:

$$\begin{aligned} b_1(x_0) &= \frac{1}{h}\int K\left(\frac{x_0-u}{h}\right) g(u)du - g(x_0) \\ &= \int K(v)(g(x_0-vh)-g(x_0))dv. \end{aligned}$$

Using the Taylor formula, there exists $\tau \in [0,1]$ such that

$$g(x_0-vh)-g(x_0) = -vhg'(x_0) + ... + \frac{(-vh)^l}{l!}g^{(l)}(x_0-\tau vh).$$

where $l = \lfloor \beta \rfloor$. Since the kernel $K$ is of order $l$,

$$\begin{aligned} b_1(x_0) &= \int K(v)\frac{(-vh)^l}{l!}g^{(l)}(x_0-\tau vh)dv \\ &= \int K(v)\frac{(-vh)^l}{l!}(g^{(l)}(x_0-\tau vh)-g^{(l)}(x_0))dv. \end{aligned}$$

From G4($\beta$), it follows that

$$|b_1(x_0)| \leq \int |K(v)|\frac{|vh|^l}{l!}L|\tau vh|^{\beta-l}dv \leq \frac{h^\beta L}{l!}\int |K(v)||v|^\beta dv$$

as $\tau \in [0,1]$ and $\int |K(v)||v|^\beta dv < +\infty$ under K1 and K2($\alpha$), $\alpha \geq \beta$.

Gathering (11) and (12), we obtain

$$\begin{aligned} b(x_0)^2 &\leq 2b_1(x_0)^2 + 2b_2(x_0)^2 \\ &\leq 2(L\int |K(v)||v|^\beta dv/l!)^2 h^{2\beta} + 2(2\|g'\|_\infty \|K\|_1\|g\|_1)^2\Delta^2. \end{aligned}$$

This completes the proof of Lemma 3.1. □

5.2. **Proof of Lemma 3.2.** We have the following relation:

$$\psi_\Delta'' = i\Delta\psi_\Delta' g^* + i\Delta\psi_\Delta g^{*\prime} = -\Delta^2\psi_\Delta {g^*}^2 + i\Delta\psi_\Delta g^{*\prime}. \tag{12}$$

As the $Z_k^\Delta$ are i.i.d., we have:

$$\mathrm{Var}[\widehat{g}(x_0)] = \mathrm{Var}\left[\frac{1}{nh\Delta}\sum_{k=1}^n Z_k^\Delta K\left(\frac{x_0-Z_k^\Delta}{h}\right)\right] = \frac{1}{n(h\Delta)^2}\mathrm{Var}\left[Z_1^\Delta K\left(\frac{x_0-Z_1^\Delta}{h}\right)\right].$$

Thus,

$$\mathrm{Var}[\widehat{g}(x_0)] \leq \frac{1}{n(h\Delta)^2}\mathbb{E}\left[(Z_1^\Delta)^2 K^2\left(\frac{x_0-Z_1^\Delta}{h}\right)\right].$$

Writing

$$K^2\left(\frac{x_0-Z_1^\Delta}{h}\right) = \left|\frac{1}{2\pi}\int K^*(u)e^{-i\frac{(x_0-Z_1^\Delta)u}{h}}du\right|^2,$$

we obtain with $v = u/h$

$$\begin{aligned}
\mathrm{Var}[\widehat{g}(x_0)] &\leq \frac{1}{n\Delta^2}\mathbb{E}\left[(Z_1^\Delta)^2\left|\frac{1}{2\pi}\int K^*(vh)e^{-i(x_0-Z_1^\Delta)v}dv\right|^2\right] \\
&\leq \frac{1}{n\Delta^2(2\pi)^2}\mathbb{E}\left[\int\int Z_1^\Delta e^{iZ_1^\Delta v}K^*(vh)e^{-ix_0v}\overline{Z_1^\Delta e^{iZ_1^\Delta u}K^*(uh)e^{-ix_0u}}dvdu\right].
\end{aligned}$$

Using Fubini and $\mathbb{E}[(Z_1^\Delta)^2 e^{iZ_1^\Delta(v-u)}] = -\psi_\Delta''(v-u)$ we find

$$\mathrm{Var}[\widehat{g}(x_0)] \leq \frac{1}{n\Delta^2(2\pi)^2}\int\int|-\psi_\Delta''(v-u)K^*(vh)K^*(uh)|dvdu.$$

Now formula (12) gives $\mathrm{Var}[\widehat{g}(x_0)] \leq T_1 + T_2$ with

$$\begin{aligned}
T_1 &= \frac{1}{n\Delta^2(2\pi)^2}\int\int|\Delta^2\psi_\Delta(v-u)(g^*)^2(v-u)K^*(vh)K^*(uh)|dvdu \\
T_2 &= \frac{1}{n\Delta^2(2\pi)^2}\int\int|\Delta\psi_\Delta(v-u)(g^*)'(v-u)K^*(vh)K^*(uh)|dvdu.
\end{aligned}$$

We first bound $T_2$:

$$\begin{aligned}
T_2 &\leq \frac{1}{n\Delta(2\pi)^2}\sqrt{\int\int|\psi_\Delta(v-u)||(g^*)'(v-u)||K^*(vh)|^2dvdu} \\
&\qquad\times\sqrt{\int\int|\psi_\Delta(v-u)||(g^*)'(v-u)||K^*(uh)|^2dvdu} \\
&\leq \frac{1}{n\Delta(2\pi)^2}\int|K^*(vh)|^2dv\int|\psi_\Delta(z)||(g^*)'(z)|dz \\
&\leq \frac{1}{nh\Delta(2\pi)^2}\int|K^*(u)|^2du\int|(g^*)'(z)|dz, \text{ setting } u = vh \text{ and because } |\psi_\Delta(z)| \leq 1 \\
&\leq \frac{\|K\|_2^2}{2\pi nh\Delta}\int|(g^*)'(z)|dz
\end{aligned}$$

where $(g^*)'$ exists and is integrable by G2. Following the same line for the study of $T_1$, we get

$$T_1 \leq \frac{\|K\|_2^2}{2\pi nh}\int|(g^*)^2(z)|dz \leq \frac{\|K\|_2^2\|g\|_2^2}{nh},$$

This completes the proof of Lemma 3.2. □

5.3. **Proof of Theorem 3.1.** The essence of the proof is to build two functions $g_0$ and $g_1$ which are far in term of pointwise distance but with close associated distribution. Let

$$g_0(x) = xf_\lambda(x) = \frac{1}{\pi}\frac{\lambda x}{1+(\lambda x)^2}$$

where $f_\lambda$ is the density of the Cauchy distribution $C(0,\lambda)$ with scale parameter $\lambda$. Here $\lambda$ is a positive and small enough real (it will be made precise later). Now let $K$ a infinitely differentiable and even function such that $\int K = 0$, $K(0) \neq 0$ and $K(x) = |x|^{-2}$ for $|x|$ large enough (say for $|x| > B$). Using this auxiliary function $K$, we can define

$$g_1(x) = g_0(x) + ch_n^\beta K\left(\frac{x - x_0}{h_n}\right)x$$

where $c$ is a constant to be specified later and

$$h_n = (n\Delta)^{-\frac{1}{2\beta+1}}.$$

We denote $N_0(x) = g_0(x)/x$ and $N_1(x) = g_1(x)/x$. Remark that if $L_{0,t} = \sum_{i=1}^{N_t} Y_i$ is a compound Poisson process with $N_t$ a Poisson process of intensity 1 and $Y_i$ Cauchy $C(0, \lambda)$ variables, then its characteristic function is

$$\psi_{0,t}(u) = \exp\left(t \int_{\mathbb{R}} (e^{iux} - 1)N_0(x)dx\right)$$

and $Z_k^{0,\Delta} = L_{0,k\Delta} - L_{0,(k-1)\Delta}$ has distribution $P_0(dx) = e^{-\Delta}\delta_0(dx) + \varphi_0(x)dx$ with

$$\varphi_0(x) = \sum_{k=1}^{\infty} e^{-\Delta}\frac{\Delta^k}{k!} f_\lambda^{*k}(x).$$

Moreover $N_1$ is a density. Indeed the definition of $K$ guarantees that $\int N_1(x)dx = \int N_0(x)dx + ch_n^\beta \int K\left(\frac{x-x_0}{h_n}\right)dx = 1$. And to ensure the positivity of $N_1$, it is sufficient to prove that $|N_1 - N_0| \leq N_0$. But, if $|x| > |x_0| + Bh_n$ ,

$$N_0^{-1}(x)|N_1(x) - N_0(x)| \leq Cch_n^{\beta+2}x^2|x - x_0|^{-2} \leq 1$$

for $c$ small enough, and if $|x| \leq |x_0| + Bh_n$,

$$N_0^{-1}(x)|N_1(x) - N_0(x)| \leq Cch_n^\beta(1 + (\lambda(|x_0| + Bh_n))^2)\|K\|_\infty \leq 1$$

for $c$ small enough. Then, if $L_{1,t} = \sum_{i=1}^{N_t} Y_i$ with $N_t$ a Poisson process of intensity 1 and $Y_i$ random variables with density $N_1$, it is a Lévy process with Lévy measure $N_1(x)dx$. We denote $\psi_{1,\Delta}$ the characteristic function of $L_{1,\Delta}$ with distribution $P_1$, and $\varphi_1$ the function such that $P_1(dx) = e^{-\Delta}\delta_0(dx) + \varphi_1(x)dx$.

Now let us denote for two probability measures $P$ and $Q$, $\chi^2(P, Q) = \int (dP/dQ - 1)^2 dQ$. In the sequel we show that

1) $g_0, g_1$ belong to $\mathcal{H}(\beta, L)$,
2) $|g_1(x_0) - g_0(x_0)| \geq C(n\Delta)^{-\frac{\beta}{2\beta+1}}$,
3) $\chi^2(P_1^n, P_0^n) \leq C < \infty$ where $P_1^n$ (resp. $P_0^n$) is the distribution of a sample $Z_1^\Delta, \ldots, Z_n^\Delta$ s.t the associated Lévy process $L_0$ (resp. $L_1$) has Lévy measure $N_0(x)dx$ (resp. $N_1(x)dx$).

Then it is sufficient to use Theorem 2.2 (see also p.80) in Tsybakov (2009) to obtain Theorem 3.1. In the following we denote all constants by $C$, even if it changes from line to line.

Proof of 1). *Belonging to the Hölder space*
To prove that our hypotheses belong to $\mathcal{H}(\beta, L)$, it is sufficient to show that, for $i = 0, 1$,

$\|g_i^{(k+1)}\|_p \le L$ where $k = \lfloor \beta \rfloor$ and $p^{-1} = 1 + k - \beta$. Indeed Hölder inequality gives

$$|g_i^{(k)}(x) - g_i^{(k)}(y)| = \left| \int g_i^{(k+1)}(v) \mathbb{1}_{[x,y]}(v) dv \right| \le \|g_i^{(k+1)}\|_p |x-y|^{\beta - k} \quad \text{for all } x, y.$$

When $x$ goes to infinity, $g_0^{(k+1)}(x) = C\lambda^{-1} x^{-k-2} + o(x^{-k-2})$ so it belongs to $\mathbb{L}^p$ since $p(k+2) = (k+2)/(k+1-\beta) > 1$. Choosing $\lambda$ small enough ensures $\|g_0^{(k+1)}\|_p \le L/2 \le L$.

Now to study $g_1$, we can write

$$(g_1 - g_0)^{(k+1)}(x) = cxK^{(k+1)}\left(\frac{x - x_0}{h_n}\right) h_n^{\beta-k-1} + c(k+1)K^{(k)}\left(\frac{x-x_0}{h_n}\right) h_n^{\beta-k}.$$

Let us see if this two terms are in $\mathbb{L}^p$. Writing $x = x - x_0 + x_0$ and changing variables

$$\int \left| xK^{(k+1)}\left(\frac{x-x_0}{h_n}\right)\right|^p dx \le 2^{p-1} h_n^{p+1} \int |vK^{(k+1)}(v)|^p dv + 2^{p-1}|x_0|^p h_n \int |K^{(k+1)}(v)|^p dv.$$

These integrals are finite since $vK^{(k+1)}(v) = v^{-(2+k)}$ for $v$ large enough and $p(k+2) = (k+2)/(k+1-\beta) > 1$. In the same way

$$\int \left| K^{(k)}\left(\frac{x-x_0}{h_n}\right)\right|^p dx \le h_n \int |K^{(k)}(v)|^p dv.$$

Thus

$$\|(g_1 - g_0)^{(k+1)}\|_p^p \le Cc^p(h_n h_n^{p(\beta-k-1)} + h_n h_n^{p(\beta-k)}) \le Cc^p h_n^{p(1/p+\beta-k-1)} \le Cc^p \le (L/2)^p$$

for suitable $c$. Then $g_1 - g_0$ belongs to $\mathcal{H}(\beta, L/2)$ and $g_1$ belongs to $\mathcal{H}(\beta, L)$.

Proof of 2). *Rate*

By assumption, $x_0 \ne 0$ and we can see that $|g_1(x_0) - g_0(x_0)| = ch_n^\beta |K(0)x_0|$ with $K(0) \ne 0$. Since $h_n = (n\Delta)^{-\frac{1}{2\beta+1}}$, this quantity has the announced order of the rate: $(n\Delta)^{-\frac{\beta}{2\beta+1}}$ .

Proof of 3). *Chi-square divergence*

Since the observations are i.i.d., $\chi^2(P_1^n, P_0^n) = (1 + \chi^2(P_1, P_0))^n - 1$. Thus, it is sufficient to prove that $\chi^2(P_1, P_0) = O(n^{-1})$ where

$$\chi^2(P_1, P_0) = \int_{x \ne 0} \left(\frac{\varphi_1(x)}{\varphi_0(x)} - 1\right)^2 \varphi_0(x) dx.$$

Indeed $P_1(\{0\}) = e^{-\Delta} = P_0(\{0\})$. Now let us remark that for $n$ large enough

$$\varphi_0(x) = \sum_{k=1}^{\infty} e^{-\Delta} \frac{\Delta^k}{k!} f_\lambda^{*k}(x) \ge e^{-\Delta} \Delta f_\lambda(x) \ge \Delta e^{-C} \lambda \pi^{-1}/(1 + (\lambda x)^2)$$

since $\Delta$ is bounded. Then $\varphi_0(x) \ge C^{-1}\Delta x^{-2}$ for $|x|$ large enough, say $|x| \ge A$ and $\varphi_0(x) \ge C^{-1}\Delta$ for $|x| \le A$. Next we write $\chi^2(P_1, P_0) = \int_{x\ne 0} (\varphi_1(x) - \varphi_0(x))^2 (\varphi_0(x))^{-1} dx = I_1 + I_2$ where $I_1$ is the integral for $|x| < A$ and $I_2$ for $|x| \ge A$. We will bound these two terms separately.

Since $\varphi_0(x) \ge C^{-1}\Delta$ for $|x|$ small

$$I_1 = \int_{|x|<A} (\varphi_1(x) - \varphi_0(x))^2 (\varphi_0(x))^{-1} dx \le C\Delta^{-1} \int_{|x|<A} (\varphi_1(x) - \varphi_0(x))^2 dx.$$

For $i = 0, 1$, the Fourier tranform of $\varphi_i$ is $\psi_{i,\Delta}(u) - P_i(\{0\})$. Thus Parseval equality gives

$$I_1 \leq C\Delta^{-1} \int |\psi_{1,\Delta}(u) - \psi_{0,\Delta}(u)|^2 \, du.$$

In order to get a bound on $|\psi_{1,\Delta} - \psi_{0,\Delta}|$, we apply the mean value theorem:

$$|\psi_1(u) - \psi_0(u)| \leq \sup_{z \in I_u} |e^z| |\Delta \int (e^{iux} - 1)(N_1(x) - N_0(x))dx|$$

where $I_u$ is the segment in $\mathbb{C}$ between $a_u = \Delta \int (e^{iux} - 1)N_0(x)dx$ and $b_u = \Delta \int (e^{iux} - 1)N_1(x)dx$. But

$$\int (e^{iux} - 1)(N_1(x) - N_0(x))dx = ch_n^\beta \int (e^{iux} - 1)K\left(\frac{x - x_0}{h_n}\right) dx = ch_n^{\beta+1} e^{iux_0} K^*(h_n u).$$

Note that this quantity is well defined since $K$ belongs to $\mathbb{L}^1$. Thus

$$|\psi_1(u) - \psi_0(u)| \leq \quad (\sup_{z \in I_u} e^{\Re(z)})\Delta ch_n^{\beta+1} |K^*(h_n u)|$$

where $\Re(x)$ means the real part of $x$. We can compute $\Re(a_u) = a_u = \Delta(N_0^*(u) - 1) = \Delta(\exp(-|u/\lambda|) - 1) \leq 0$ and

$$\Re(b_u) = \Re(\Delta(N_0^*(u) - 1 + (N_1 - N_0)^*(u))) = \Delta(N_0^*(u) - 1 + ch_n^{\beta+1}\Re(K^*(h_n u)e^{iux_0})).$$

Since $K$ is even,

$$\Re(b_u) = \Delta(\exp(-|u/\lambda|) - 1 + ch_n^{\beta+1} K^*(h_n u)\cos(ux_0)) \leq c\Delta h_n^{\beta+1} \|K^*\|_\infty \leq C$$

so that

$$|\psi_1(u) - \psi_0(u)| \leq \quad e^C \Delta ch_n^{\beta+1} |K^*(h_n u)|. \tag{13}$$

Then

$$I_1 \leq C\Delta^{-1} \int \left|\Delta h_n^{\beta+1} K^*(h_n u)\right|^2 du \leq C\Delta h_n^{2\beta+1}. \tag{14}$$

Let us now bound the term $I_2$, using that $\varphi_0(x) \geq C^{-1}\Delta x^{-2}$ for $|x|$ large enough

$$I_2 \quad = \quad \int_{|x| \geq A} \frac{(\varphi_1(x) - \varphi_0(x))^2}{\varphi_0(x)} dx \leq C\Delta^{-1} \int (\varphi_1(x) - \varphi_0(x))^2 \, x^2 dx.$$

But $F = \varphi_1 - \varphi_0$ has Fourier transform

$$F^* = \psi_{1,\Delta} - \psi_{0,\Delta} = \exp(\Delta(e^{-|u/\lambda|} + ch_n^{\beta+1} K^*(h_n u)e^{iux_0} - 1)) - \exp(\Delta(e^{-|u/\lambda|} - 1))$$

and this function is differentiable everywhere exept at $u = 0$, with derivative

$$F^{*\prime} = \Delta\gamma_1\psi_{1,\Delta} - \Delta\gamma_0\psi_{0,\Delta}$$

where

$$\gamma_0(u) = -\text{sign}(u).e^{-|u/\lambda|}/\lambda, \qquad \gamma_1(u) = \gamma_0(u) + ch_n^{\beta+1} e^{iux_0}(ix_0 K^*(h_n u) + h_n K^{*\prime}(h_n u)).$$

Let us now prove that the Fourier transform of $F^{*\prime}$ is $-2\pi i x F(-x)$. Let us write the factorization

$$\Delta^{-1} F^{*\prime} = \gamma_1\psi_{1,\Delta} - \gamma_0\psi_{0,\Delta} = (\gamma_1 - \gamma_0)\psi_{1,\Delta} + \gamma_0(\psi_{1,\Delta} - \psi_{0,\Delta}) \tag{15}$$

with $|\psi_{1,\Delta}| \leq 1$. Since $K^*$ and $K^{*\prime}$ are uniformly bounded, $\gamma_1 - \gamma_0$ is bounded as well. In the same way, the inequality (13) entails that $\|\psi_{1,\Delta} - \psi_{0,\Delta}\|_\infty < \infty$, so that $F^{*\prime}$ is bounded. Thus $F^*$ is Lipschitz and absolutely continuous. Moreover, using again (15), we can see that $F^{*\prime}$ is integrable (we can choose $K$ such that $K^*$ is integrable, for example take for $K$ the difference between the Cauchy density and the normal density). Then, according to Rudin (1987), the Fourier transform of $F^{*\prime}$ is $-ixF^{**}(x)$ (it is in fact a simple integration by parts). Since $F^*$ is integrable, $F^{**}(x) = 2\pi F(-x)$ almost everywhere, and we have proved that $(F^{*\prime})^*(x) = -2\pi ixF(-x)$ a.e.. Next, the Parseval equality provides $\int |xF(x)|^2dx = (2\pi)^{-1}\int |F^{*\prime}(u)|^2du$. Thus

$$I_2 \leq C\Delta^{-1}\int |xF(x)|^2dx \leq C\Delta(2\pi)^{-1}\int |\gamma_1\psi_{1,\Delta} - \gamma_0\psi_{0,\Delta}|^2.$$

Hence, using the factorization (15) we can split $I_2 \leq \pi^{-1}C\Delta(I_{2,1} + I_{2,2})$ with

$$\begin{cases} I_{2,1} = \int |\gamma_1 - \gamma_0|^2, \\ I_{2,2} = \int |\gamma_0(\psi_{1,\Delta} - \psi_{0,\Delta})|^2. \end{cases}$$

Using the definition of $\gamma_1$, we compute

$$\begin{aligned} I_{2,1} &= c^2h_n^{2\beta+2}\int |ix_0K^*(h_nu) + h_nK^{*\prime}(h_nu)|^2du \\ &\leq 2c^2h_n^{2\beta+1}\left(x_0^2\int |K^*|^2 + h_n^2\int |K^{*\prime}|^2\right) \\ &\leq 4\pi c^2h_n^{2\beta+1}\left(x_0^2\int |K|^2 + h_n^2\int |xK(x)|^2\right) \leq Ch_n^{2\beta+1}. \end{aligned} \tag{16}$$

Now, in order to deal with $I_{2,2}$, we use the previous bound (13) on $|\psi_{1,\Delta} - \psi_{0,\Delta}|$

$$\begin{aligned} I_{2,2} &\leq Cc^2\Delta^2h_n^{2\beta+2}\int |\gamma_0(u)K^*(h_nu)|^2du \\ &\leq Cc^2\Delta^2h_n^{2\beta+2}\|K^*\|_\infty\|\gamma_0\|_2^2 \leq Ch_n^{2\beta+1} \end{aligned} \tag{17}$$

since $\Delta$ is bounded.

Finally, by gathering (14), (16) and (17), we get

$$\chi^2(P_1, P_0) \leq C\Delta h_n^{2\beta+1} = O(n^{-1}).$$

This ends the proof of Theorem 3.1. □

5.4. **Proof of Theorem 3.2.** The goal is to bound $\mathbb{E}[|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2]$. To do this, we fix $h \in H$. We write

$$|g(x_0) - \hat{g}_{\hat{h}}(x_0)| \leq |\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)| + |\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)| + |\hat{g}_h(x_0) - g(x_0)|.$$

So we have

$$|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2 \leq 3|\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)|^2 + 3|\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)|^2 + 3|\hat{g}_h(x_0) - g(x_0)|^2.$$

Define $B := |\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)|^2$ and $C := |\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)|^2$.

We have $A(h) \geq |\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)|^2 - V(\hat{h}) \geq B - V(\hat{h})$. So $B \leq A(h) + V(\hat{h})$.

Moreover, $A(\hat{h}) \geq |\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)|^2 - V(h) \geq C - V(h)$. So $C \leq A(\hat{h}) + V(h)$.

Therefore,

$$|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2 \leq 3(A(h) + V(\hat{h})) + 3(A(\hat{h}) + V(h)) + 3|\hat{g}_h(x_0) - g(x_0)|^2.$$

Now, by definition of $\hat{h}$, $A(\hat{h}) + V(\hat{h}) \le A(h) + V(h)$. This allows us to write

$$|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2 \le 6A(h) + 6V(h) + 3|\hat{g}_h(x_0) - g(x_0)|^2.$$

Let us denote $b_h(x_0) = \mathbb{E}[\hat{g}_h(x_0)] - g(x_0)$ and $b_{h,2}(x_0) = \mathbb{E}[\hat{g}_h(x_0)] - K_h \star g(x_0)$ (these are the same notations as in Lemma 3.1, but with subscript $h$). Thus

$$\begin{aligned} \mathbb{E}[|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2] &\le 6\mathbb{E}[A(h)] + 6V(h) + 3b_h^2(x_0) + 3\mathrm{Var}(\hat{g}_h(x_0)) \\ &\le 6\mathbb{E}[A(h)] + 3b_h^2(x_0) + C_2 V(h). \end{aligned}$$

It remains to bound $\mathbb{E}[A(h)]$. Let us denote by $g_{h,h'} = \mathbb{E}[\hat{g}_{h,h'}]$ and $g_h = \mathbb{E}[\hat{g}_h]$. We write

$$\hat{g}_{h,h'} - \hat{g}_{h'} = \hat{g}_{h,h'} - g_{h,h'} - \hat{g}_{h'} + g_{h'} + g_{h,h'} - g_{h'}, \tag{18}$$

and we study the last term of the above decomposition. We have

$$\begin{aligned} |g_{h,h'}(x_0) - g_{h'}(x_0)| &= |\mathbb{E}[\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0)]| \\ &= |\mathbb{E}[K_{h'} \star \hat{g}_h(x_0) - \hat{g}_{h'}(x_0)]| \\ &= |K_{h'} \star \mathbb{E}[\hat{g}_h(x_0) - g(x_0)] + K_{h'} \star g(x_0) - \mathbb{E}[\hat{g}_{h'}(x_0)]|. \end{aligned}$$

This can be written:

$$\begin{aligned} |g_{h,h'}(x_0) - g_{h'}(x_0)| &= |K_{h'} \star b_h(x_0) + b_{h,2}(x_0)| \\ &\le \left| \int K\left(\frac{x_0 - u}{h'}\right) b_h(u) \frac{du}{h'} \right| + |b_{h,2}(x_0)|. \end{aligned}$$

Now $|b_{h,2}(x_0)| \le |b_h(x_0)| \le \|b_h\|_\infty$ so that

$$\begin{aligned} |g_{h,h'}(x_0) - g_{h'}(x_0)|^2 &\le 2\|b_h\|_\infty^2 \left( \int |K(v)| dv \right)^2 + 2|b_{h,2}(x_0)|^2 \\ &\le 2(\|K\|_1^2 + 1)\|b_h\|_\infty^2. \end{aligned} \tag{19}$$

Then by inserting (19) in decomposition (18), we find:

$$\begin{aligned} A(h) &= \sup_{h'} \{|\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0)|^2 - V(h')\}_+ \\ &\le 3\sup_{h'} \{|\hat{g}_{h,h'}(x_0) - g_{h,h'}(x_0)|^2 - V(h')/6\}_+ \\ &\quad +3\sup_{h'} \{|\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - V(h')/6\}_+ + 6(\|K\|_1^2 + 1)\|b_h\|_\infty^2. \end{aligned} \tag{20}$$

We can prove the following concentration result:

**Proposition 5.3.** *Assume that $g$ satisfies G1, G2, G3(5) , $K$ satisfies K1, $M = O((n\Delta)^{1/3})$ and take $c$ in (9) such that $c \ge 16\max(1, \|K\|_\infty)$. Then*

$$\mathbb{E}\left[\sup_{h'} \{|\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - V(h')/6\}_+\right] = O\left(\frac{\log(n\Delta)}{n\Delta}\right) \tag{21}$$

$$\mathbb{E}\left[\sup_{h'} \{|\hat{g}_{h,h'}(x_0) - g_{h,h'}(x_0)|^2 - V(h')/6\}_+\right] = O\left(\frac{\log(n\Delta)}{n\Delta}\right). \tag{22}$$

Inequalities (21) et (22) together with (20) imply

$$\mathbb{E}[|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2] \le C_1\|b_h\|_\infty^2 + C_2V(h) + C_3\frac{\log(n\Delta)}{n\Delta}.$$

This completes the proof of Theorem 3.2. $\square$

5.5. **Proof of Proposition 5.3.** Note that

$$\hat{g}_{h'}(x_0) - g_{h'}(x_0) = \frac{1}{n}\sum_{k=1}^{n}\left[\frac{{Z_k}^\Delta}{\Delta}K_{h'}\left(x_0 - {Z_k}^\Delta\right) - \mathbb{E}\left(\frac{{Z_k}^\Delta}{\Delta}K_{h'}\left(x_0 - {Z_k}^\Delta\right)\right)\right]. \tag{23}$$

In order to apply a Bernstein inequality, since the $Z_k^\Delta$'s are not bounded, we truncate these variables and consider the following decomposition:

$$\{|{Z_k}^\Delta| \le \mu_n\} \text{ and } \{|{Z_k}^\Delta| > \mu_n\}$$

where

$$\mu_n = \mu_n(h') = \frac{\|K\|_2^2(\|(g^*)'\|_1 + \Delta\|g\|_2^2)}{2\pi\|K\|_\infty\sqrt{V(h')/6}}. \tag{24}$$

We decompose then (23) as follows

$$\begin{aligned}\hat{g}_{h'}(x_0) - g_{h'}(x_0) &= \frac{1}{n}\sum_{k=1}^{n} W_k(h') + T_k(h') - \mathbb{E}\left(W_k(h') + T_k(h')\right)\\ &= S_n(W(h')) + S_n(T(h'))\end{aligned}$$

where $S_n(X)$ means $(1/n)\sum_{i=1}^n[X_i - \mathbb{E}(X_i)]$ and

$$W_k(h) = \frac{{Z_k}^\Delta}{\Delta}K_h\left(x_0 - {Z_k}^\Delta\right)\mathbb{1}_{\{|{Z_k}^\Delta| \le \mu_n(h)\}} \tag{25}$$

$$T_k(h) = \frac{{Z_k}^\Delta}{\Delta}K_h\left(x_0 - {Z_k}^\Delta\right)\mathbb{1}_{\{|{Z_k}^\Delta| > \mu_n(h)\}}. \tag{26}$$

Thus

$$\begin{aligned}&\mathbb{E}\left[\sup_{h'}\{|\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - V(h')/6\}_+\right]\\ \le\ & 2\sum_{h'\in H}\mathbb{E}\left[S_n(W(h'))^2 - V(h')/12\right]_+ + 2\sum_{h'\in H}\mathbb{E}\left[S_n(T(h')^2\right].\end{aligned}$$

Then we use the two following lemmas

**Lemma 5.1.** *Assume that $g$ satisfies G1, G2, $K$ satisfies K1, and $c \ge 16, M = O((n\Delta)^{1/3})$. Then there exists $C > 0$ only depending on $K$ and $g$ such that*

$$\sum_{h\in H}\mathbb{E}\left[S_n^2(W(h)) - V(h)/12\right]_+ \le C\frac{\log(n\Delta)}{n\Delta}.$$

**Lemma 5.2.** *Under assumptions K1, G3(5) and if $M = O((n\Delta)^{1/3})$,*

$$\sum_{h\in H}\mathbb{E}\left[S_n^2(T(h))\right] \le C'\frac{1}{n\Delta}.$$

Lemmas 5.1 and 5.2 yield

$$\mathbb{E}\left[\sup_{h'}\{|\hat{g}_{h'}(x_0)-g_{h'}(x_0)|^2-V(h')/6\}_+\right]\le C''\left(\frac{1}{n\Delta}+\frac{\log(n\Delta)}{n\Delta}\right)$$

Inegality (22) is obtained by following the same lines as for inequality (21) with $K_h$ replaced by $K_{h'}\star K_h$. This ends the proof of Proposition 5.3. □

5.6. **Proof of lemma 5.1.** First, note that

$$\begin{aligned}\mathbb{E}\left[S_n^2(W(h))-V(h)/12\right]_+ &\le \int_0^\infty \mathbb{P}(S_n^2(W(h))\ge V(h)/12+x)dx\\ &\le \int_0^\infty V(h)\mathbb{P}\left(|S_n(W(h))|\ge\sqrt{V(h)(1/12+y)}\right)dy.\end{aligned}$$

Next, we recall the classical Bernstein inequality (see e.g. Birgé and Massart (1998) for a proof):

**Lemma 5.3.** *Let $W_1,...,W_n$ $n$ independent and identically distributed random variables and $S_n(W)=(1/n)\sum_{i=1}^n[W_i-\mathbb{E}(W_i)]$. Then, for $\eta>0$,*

$$\mathbb{P}(|S_n(W)|\ge\eta)\quad\le\quad 2\exp\left(\frac{-n\eta^2/2}{\nu^2+b\eta}\right)\le 2\max\left(\exp\left(\frac{-n\eta^2}{4\nu^2}\right),\exp\left(\frac{-n\eta}{4b}\right)\right),$$

*where $Var(W_1)\le\nu^2$ and $|W_1|\le b$.*

We apply this form of Bernstein inequality to $W_i(h)$ defined by (25) and $\eta=\sqrt{(1/12+y)V(h)}$. Using Lemma 3.2, it is easy to see that

$$\mathrm{Var}(W_i)\le\nu^2:=\frac{\|K\|_2^2(\|(g^*)'\|_1+\Delta\|g\|_2^2)}{2\pi\Delta h}\text{ and }|W_i|\le b:=\frac{\|K\|_\infty\mu_n(h)}{\Delta h}.$$

We find

$$\begin{aligned}\exp\left(\frac{-n\eta^2}{4\nu^2}\right) &= \exp\left(-\frac{\pi(1/12)V(h)n\Delta h}{2\|K\|_2^2(\|(g^*)'\|_1+\Delta\|g\|_2^2)}\right)\times\exp\left(-\frac{\pi yV(h)n\Delta h}{2\|K\|_2^2(\|(g^*)'\|_1+\|g\|_2^2)}\right)\\ &= (n\Delta)^{-c/48}\times(n\Delta)^{-cy/4}\end{aligned}$$

and

$$\exp\left(\frac{-n\eta}{4b}\right)\quad\le\quad(n\Delta)^{-c/48}\times(n\Delta)^{-c\sqrt{y/192}}.$$

Then we deduce

$$\begin{aligned}\mathbb{E}\left[S_n^2(W(h))-V(h)/12\right]_+ &\le \int_0^\infty V(h)(n\Delta)^{-c/48}\max\left((n\Delta)^{-cy/4},(n\Delta)^{-c\sqrt{y/192}}\right)dy\\ &\le V(h)(n\Delta)^{-c/48}\left(\int_0^\infty(n\Delta)^{-cy/4}dy+\int_0^\infty(n\Delta)^{-c\sqrt{y/192}}dy\right)\\ &\le \frac{4}{c}V(h)(n\Delta)^{-c/48}\left(\frac{1}{\log(n\Delta)}+\frac{96}{c\log(n\Delta)^2}\right)\end{aligned}$$

using that $\int_0^\infty e^{-y/\lambda} = \lambda$ and $\int_0^\infty e^{-\sqrt{y}/\lambda} = 2\lambda^2$. Replacing $V(h)$ by its value, it gives

$$\sum_{h\in H} \mathbb{E}\left[S_n^2(W(h)) - V(h)/12\right]_+ \leq \frac{4C_0}{c}(n\Delta)^{-1-c/48}\left(1+\frac{96}{c\log(n\Delta)}\right)\sum_{h\in H}\frac{1}{h}.$$

Recall that $H = \{\frac{k}{M}, 1 \leq k \leq M\}$. Then

$$\sum_h \frac{1}{h} = \sum_{k=1}^{M}\frac{M}{k} \leq \log(M)M \leq \frac{1}{3}\log(n\Delta)(n\Delta)^{1/3}.$$

Finally

$$\begin{aligned}\sum_{h\in H} \mathbb{E}\left[S_n^2(W(h)) - V(h)/12\right]_+ &\leq \frac{4C_0}{3c}(n\Delta)^{-2/3-c/48}\left(\log(n\Delta)+\frac{96}{c}\right)\\ &\leq \frac{4C_0}{3c}(n\Delta)^{-1}\left(\log(n\Delta)+\frac{96}{c}\right)\end{aligned}$$

as soon as $c \geq 16$. This completes the proof of lemma 5.1. □

5.7. **Proof of lemma 5.2.** For a fixed bandwidth $h$ in $H$, we can establish the following bound:

$$\begin{aligned}\mathbb{E}\left[|S_n(T(h))|^2\right] &= \mathrm{Var}\left[\frac{1}{n}\sum_{k=1}^{n}\frac{Z_k^\Delta}{\Delta h}K\left(\frac{x_0-Z_k^\Delta}{h}\right)\mathbb{1}_{\{|Z_k^\Delta|>\mu_n\}}\right]\\ &\leq \frac{1}{n}\frac{\|K\|_\infty^2}{(\Delta h)^2}\mathbb{E}[(Z_1^\Delta)^2\mathbb{1}_{\{|Z_1^\Delta|>\mu_n\}}]\\ &\leq \frac{1}{n\Delta}\frac{\|K\|_\infty^2}{h^2}\frac{\mathbb{E}[|Z_1^\Delta|^{w+2}/\Delta]}{{\mu_n}^w}\end{aligned}$$

for any $w > 0$. Recall that, according to Proposition 5.2, $\mathbb{E}[|Z_1^\Delta|^{w+2}/\Delta]$ is bounded under G3$(w+2)$. We search conditions for $\sum_h h^{-2}{\mu_n}^{-w} \leq$ constant. The following equalities hold up to constants:

$$\sum_{h\in H}\frac{1}{h^2{\mu_n}^w} = \sum_h \frac{V(h)^{w/2}}{h^2} = \frac{\log(n\Delta)^{w/2}}{(n\Delta)^{w/2}}\sum_h\frac{1}{h^{2+w/2}}$$

Since $h = k/M$, this provides

$$\sum_h \frac{1}{h^{2+w/2}} = \sum_{k=1}^{M}\left(\frac{M}{k}\right)^{2+w/2} = M^{2+w/2}\sum_{k=1}^{M}\frac{1}{k^{2+w/2}} = O(M^{2+w/2}).$$

Finally, as $M = O((n\Delta)^{1/3})$, we have

$$\sum_h \frac{1}{h^2{\mu_n}^w} \leq C\frac{M^{2+w/2}\log(n\Delta)^{w/2}}{(n\Delta)^{w/2}} \leq C\log(n\Delta)^{w/2}(n\Delta)^{\frac{1}{3}(2+\frac{w}{2})-\frac{w}{2}}$$

We need that $(2+w/2)\times 1/3 - w/2 < 0$, so we need the $Z_i$ admit a moment of order $w+2 \geq 5$. This completes the proof of lemma 5.2. □

## Acknowledgement

The authors thank Valentine Genon-Catalot and Fabienne Comte for enlightening discussions and helpful advices.


## References


Belomestny, D. (2011). Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.*, 39(4):2205–2242.

Bertoin, J. (1996). *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York.

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375.

Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.*, 5:1–31 (electronic).

Comte, F. and Genon-Catalot, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Process. Appl.*, 119(12):4088–4123.

Comte, F. and Genon-Catalot, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(3):595–617.

Duval, C. (2012). Adaptive wavelet estimation of a compound Poisson process. *arXiv:1203.3135*.

Figueroa-López, J. E. (2011). Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, 17(2):643–670.

Figueroa-López, J. E. and Houdré, C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 96–116. Inst. Math. Statist., Beachwood, OH.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632.

Jongbloed, G. and van der Meulen, F. H. (2006). Parametric estimation for subordinators and induced OU processes. *Scand. J. Statist.*, 33(4):825–847.

Kerkyacharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields*, 121:137–170.

Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition.

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

van Es, B., Gugushvili, S., and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694.

Watteel, R. N. and Kulperger, R. J. (2003). Nonparametric estimation of the canonical measure for infinitely divisible distributions. *J. Stat. Comput. Simul.*, 73(7):525–542.