



# Adaptive pointwise estimation for pure jump Lévy processes

Mélina Bec, Claire Lacour

## ► To cite this version:

Mélina Bec, Claire Lacour. Adaptive pointwise estimation for pure jump Lévy processes. *Statistical Inference for Stochastic Processes*, 2015, 18 (3), pp.229-256. 10.1007/s11203-014-9113-6 . hal-00583221v4

**HAL Id: hal-00583221**

**<https://hal.science/hal-00583221v4>**

Submitted on 12 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ADAPTIVE POINTWISE ESTIMATION FOR PURE JUMP LÉVY PROCESSES

MÉLINA BEC\*, CLAIRE LACOUR\*\*

ABSTRACT. This paper is concerned with adaptive kernel estimation of the Lévy density  $N(x)$  for bounded-variation 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  $n\Delta$  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. We also consider the case of irregular sampling.

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

March 9, 2015

## 1. INTRODUCTION

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

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

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  $(L_{k\Delta}, k = 1, \dots, n)$  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  $(Z_k^{\Delta} = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \dots, n)$  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  $\nu(x) = x^2N(x)$  in the more general case with drift and volatility, and Comte and Genon-Catalot (2010b) use model selection to build an adaptive estimator. An adaptive method to estimate linear functionals

---

\* UMR CNRS 8145 MAP5, Université Paris Descartes, \*\* Laboratoire de Mathématiques d'Orsay, Université Paris-Sud.

is also given in Kappus (2012). Belomestny (2011) addresses the issue of inference for time-changed Lévy processes with results in term of uniform and pointwise distance. One can also cite Gugushvili (2012) or Nickl and Reiß (2012) for recent works at fixed  $\Delta$ .

In the high frequency context, which is our concern in this paper, the problem is simpler since, for any fixed  $u$ ,  $\psi_\Delta(u) \rightarrow 1$  when  $\Delta \rightarrow 0$ . 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:

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

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:

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

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 measure:

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

weakly converges to  $g(z)dz$  (note that the idea of exploiting this weak convergence is already present in Figueroa-López (2009b)). This suggests to consider kernel estimators of  $g$  of the form

$$(3) \quad \hat{g}_h(x) = K_h \star \hat{\mu}_n(x) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k^\Delta K_h(x - Z_k^\Delta)$$

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, but for integrated distance, we can cite Ueltzhöfer and Klüppelberg (2011), and Duval (2012) for the estimation of a compound Poisson process with low conditions on  $\Delta$ . Bücher and Vetter (2013) deal with the multivariate case.

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 developed by Goldenshluger and Lepski (2011) for density estimation. The advantage of our kernel method is that it allows us to estimate the Lévy density at a fixed point, with a local adaptive choice. This method is easy to implement, and we show its good numerical performance on different examples. Moreover our contribution includes an alternative proof for a lower bound result (see Figueroa-López (2009a)) which proves the optimality of the rate for this pointwise estimation. We also study the framework of irregular sampling.

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 (3) 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  $(\log(n\Delta)/n\Delta)^{\beta/(2\beta+1)}$  which is expected in adaptive pointwise context. Examples and simulations in our framework are discussed in Section 4. The case of irregular sampling is addressed in Section 5 and proofs are gathered in Section 6.

## 2. NOTATIONS AND ASSUMPTIONS

We present the assumptions on the kernel  $K$  and on the function  $g$  required to study the estimator given by (3). 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  $\|u\|_2$ ,  $\langle u, v \rangle$ ,  $u \star v$  the quantities

$$\|u\|_2 = \left( \int |u(x)|^2 dx \right)^{1/2},$$

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

We shall also use  $\|u\|_1 = \int |u(x)| dx$  and  $\|u\|_\infty = \sup_{x \in \mathbb{R}} |u(x)|$ . For a positive real  $\beta$ ,  $[\beta]$  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  $l = [\beta]$ . The Hölder class  $\mathcal{H}(\beta, L)$  on  $\mathbb{R}$  is the set of all functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  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  $g$ , defined by  $g(x) = xN(x)$ , where  $N$  is the Lévy density.

**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  $\mathbb{E}|Z_1^\Delta|^p < \infty$ . G4 is a classical regularity assumption in nonparametric estimation; it allows to quantify the bias (see Tsybakov (2009)). Note that G5 and G3(2) imply G1. Moreover G5 implies that  $g \in \mathcal{H}(1, L')$  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  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a kernel of order  $m$  if functions  $u \mapsto u^j K(u)$ ,  $j = 0, 1, \dots, m$  are integrable and satisfy

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

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  $\lfloor \beta \rfloor$  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  $l$ ,

$$(4) \quad K(t) = \sum_{k=1}^l \binom{l}{k} (-1)^{k+1} \frac{1}{k} u\left(\frac{t}{k}\right).$$

The kernel  $K$  defined by (4) is a kernel of order  $l$  which also satisfies K1 (see Kerkycharian et al. (2001) and Goldenshluger and Lepski (2011)). As usual, we define  $K_h$  by

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

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  $\hat{g} = \hat{g}_h$ , defined in (3). The usual bias variance decomposition of the Mean Squared Error yields:

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

But the bias needs further decomposition:

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

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,

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

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( $\beta$ )*

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

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

Moreover, the variance is controlled as follows:

**Lemma 3.2.** *Under G1, G2, G3(2) and if the kernel satisfies K1, we have*

$$\text{Var}[\hat{g}(x_0)] \leq c_2 \frac{1}{nh\Delta} + c'_2 \frac{1}{nh}$$

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

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

**Proposition 3.1.** *Under G2, G3(2), G4(β), G5 and if K satisfies K1 and K2(β), we have*

$$(5) \quad \text{MSE}(x_0, h) \leq c_1 h^{2\beta} + c_2 \frac{1}{nh\Delta} + c'_2 \frac{1}{nh} + c'_1 \Delta^2.$$

Recall that  $\Delta = \Delta(n)$  is such that  $\lim_{n \rightarrow +\infty} \Delta = 0$ , thus  $1/nh$  is negligible compared to  $1/nh\Delta$ . For the two first terms the optimal choice of  $h$  is  $h_{\text{opt}} \propto ((n\Delta)^{-\frac{1}{2\beta+1}})$  and the associated rate has classical 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

$$(C^*) \quad \Delta = O(n^{-1/3}).$$

**Proposition 3.2.** *Under the assumptions of Proposition 3.1 and under condition (C\*), the choice  $h_{\text{opt}} \propto ((n\Delta)^{-\frac{1}{2\beta+1}})$  minimizes the risk bound (5) and gives  $\text{MSE}(x_0, h_{\text{opt}}) = O((n\Delta)^{-\frac{2\beta}{2\beta+1}})$ . As a consequence  $\mathbb{E}[(\hat{g}(x_0)/x_0 - N(x_0))^2] = O((n\Delta)^{-\frac{2\beta}{2\beta+1}})$ .*

We can link this result to the one of Figueroa-López (2011) who proves that his projection estimator  $\hat{N}$  is such that  $(\hat{N}(x_0) - N(x_0))(n\Delta)^\alpha$  tends to a normal distribution for any  $0 < \alpha < \beta/(2\beta + 1)$ . Note that when condition (C\*) is not satisfied, the rate of convergence is spoiled. For example, if  $\Delta \sim n^{-\gamma}$  with  $0 < \gamma < 1$ , then the rate of convergence becomes  $(n\Delta)^{-2 \min(\frac{\beta}{2\beta+1}, \frac{\gamma}{1-\gamma})}$ .

The rate obtained in Proposition 3.2 turns out to be the optimal minimax rate of convergence over the class  $\mathcal{H}(\beta, L)$ . This result is proved in Figueroa-López (2009a) in the more general case of estimators based on the whole path of the process up to time  $n\Delta$ . In our case of discrete sampling, another proof is given in Section 6.3, where we prove the following result:

**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  $Z_1^\Delta, \dots, Z_n^\Delta$ , and for  $n$  large enough,*

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

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

**3.2. 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 bandwidths are possible. We set:

$$(6) \quad V(h) = C_0 \frac{\log(n\Delta)}{nh\Delta}$$

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, as an estimator of the bias,

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

Heuristically, this term has the same order as  $\sup_{h' \in H} [\mathbb{E}(\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0))]^2$  because the distance to the expectation is canceled by  $V(h')$ . And, if  $h'$  tends to 0,  $\mathbb{E}(\hat{g}_{h,h'}(x_0)) - \mathbb{E}(\hat{g}_{h'}(x_0))$  tends to  $\mathbb{E}(\hat{g}_h(x_0)) - g(x_0)$ . The precise link with the bias is detailed in the proofs. Then, the adaptive bandwidth  $h$  is chosen as follows:

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

This can be seen as a bias-variance trade-off since  $V(h)$  is close to the variance.

Before to study the performance of our final estimator  $\hat{g}_{\hat{h}}(x_0)$ , let us clarify the observation context. We still work in the high frequency framework, and we have seen that we need condition (C\*). Thus, the assumption on the observation step is the following

**S:**  $\Delta \rightarrow 0$  and  $n\Delta \rightarrow \infty$ . Moreover  $\Delta \leq 1$  and  $\Delta = O(n^{-1/3})$

We can now state the following oracle inequality.

**Theorem 3.2.** *We use a kernel satisfying K1 and a set of bandwidth  $H = \{\frac{j}{M}, 1 \leq j \leq M\}$  with  $M = O((n\Delta)^{1/3})$ . Assume that  $g$  satisfies G1, G2, G3(5) and take  $V(h)$  such that*

$$(7) \quad C_0 = C_0(c) = \frac{c}{2\pi} \|K\|_2^2 (\|(g^*)'\|_1 + \|g^*\|_2^2)$$

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

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

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 pointwise control of the bias has been replaced by an uniform control. Actually, it is possible to keep the pointwise risk in the right term at the cost of a supplementary term  $\sup_{h' \in H} |K_{h'} \star b_h(x_0)|^2$ . Although our estimator is not linear (we have an extra bias), it is exactly the same situation as in Goldenshluger and Lepski (2013), and we can conjecture it is in some sense unavoidable.

Note that the theorem is valid for  $c$  large enough, say  $c \geq c_0$ . In the proof, we obtain the upper bound  $16 \max(1, \|K\|_\infty)$  for  $c_0$ , unfortunately we can conjecture that this bound is not the optimal one. To obtain a sharper bound we have tuned  $c_0$  in the simulation study.

The definition of the estimator uses  $\|(g^*)'\|_1$  and  $\|g^*\|_2^2$ , but these quantities can be estimated with a preliminar estimator of  $g^*$ . More precisely, we set  $K_0^* = \mathbb{1}_{[-1,1]}$  and

$$\|\widehat{(g^*)'}\|_1 = \int \left| \frac{1}{n\Delta} \sum_{k=1}^n (Z_k^\Delta)^2 K_0^*(uh_1) e^{iuZ_k^\Delta} \right| du \quad \text{with } h_1 = (n\Delta)^{-1/3},$$

$$\widehat{\|g^*\|_2^2} = \|\hat{g}_{h_2}^*\|_2^2 = \int \left| \frac{1}{n\Delta} \sum_{k=1}^n Z_k^\Delta K_0^*(uh_2) e^{iuZ_k^\Delta} \right|^2 du \quad \text{with } h_2 = (n\Delta)^{-1/3}.$$

We introduce the following smoothness condition: a function  $\psi$  belongs to the Sobolev space  $\text{Sob}(1)$  if  $\int |\psi^*(u)|^2 |u|^2 du < \infty$  (this means that  $\psi$  has a derivative which is square-integrable). Then, reinforcing the conditions on  $g$ , we obtain a similar theorem with an empirical  $C_0$ .

**Theorem 3.3.** *We use a kernel satisfying K1 and K2(1) and  $M = O((n\Delta)^{1/3})$ . Assume that  $g$  satisfies G2, G3(32), G4(1), G5. Assume also that  $g$  and  $xg(x)$  belong to  $\text{Sob}(1)$ . In the definition of  $\hat{h}$ , replace  $V(h)$  by  $\hat{V}(h) = \hat{C}_0 \log(n\Delta)/(nh\Delta)$  where*

$$\hat{C}_0 = \frac{c}{2\pi} \|K\|^2 \left( \|\widehat{(g^*)'}\|_1 + \widehat{\|g^*\|_2^2} \right)$$

with  $c \geq 32 \max(1, \|K\|_\infty)$ . Then, under scheme S,

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

where  $V(h)$  is defined by (6) and (7).

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  $h_{\text{opt}} \propto (\log(n\Delta)/n\Delta)^{1/(2\beta+1)} \geq (n\Delta)^{-1/3}$  belongs to  $H$  as soon as  $M$  is larger than a constant times  $(n\Delta)^{1/3}$ . Hence we can state the following corollary.

**Corollary 3.1.** *Assume that  $g$  satisfies G2, G3(5), G4( $\beta$ ) with  $\beta \geq 1$  and G5. We choose a kernel satisfying K1 and K2( $\beta$ ), and  $M = \lfloor (n\Delta)^{1/3} \rfloor$ . Take  $C_0$  as in Theorem 3.2 (or as in Theorem 3.3 with assumptions of this latter theorem). Then, under scheme S,*

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

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 four different processes (listed in Examples 1-4 below). As usual in nonparametric estimation, to obtain the rate of convergence, the kernel has to be of order larger than  $\beta$ , or, equivalently, the smoothness has to be smaller than the order of the kernel. In practice this does not play a big role, so we use the kernel described in (4) with  $l = 2$  and  $u$  the Gaussian density.

The bandwidth set has been fixed to  $H = \{\frac{j}{2M}, 1 \leq j \leq M\}$  with  $M = \lfloor 2(n\Delta)^{1/3} \rfloor$ . For the implementation, a difficulty is the proper calibration of the constant  $c$  in (7). This is usually done by a large number of preliminary simulations. We have chosen  $c = 0.1$  as the adequate value for a variety of models and number of observations (as previously announced, this practical  $c$  is different from the theoretical one). The estimation and adaptation are done for 50 points  $x_0$  on the abscissa interval. For clarity, we have computed the Mean Integrated Square Error (MISE) of the estimators. We also give the  $MSE(x_0) = MSE(x_0, \hat{h})$  in some points  $x_0$ , by way of example. Illustratively, Figures 1 and 2 plot



ten estimated curves corresponding to our four examples with in the first column  $\Delta = 0.2, n = 5.10^3$ , and in the second  $\Delta = 0.05, n = 5.10^4$ .

**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

$$(8) \quad \psi_t(u) = \exp \left( \lambda t \int_{\mathbb{R}} (e^{iux} - 1) f(x) dx \right).$$

Its Lévy density is  $N(x) = \lambda f(x)$  and thus  $g(x) = \lambda x f(x)$ . For our first example, we choose  $\lambda = 2$  and  $f$  such that  $g(x) = x f(x) = (1/2)\sqrt{x/2}$  for  $0 < x \leq 2$ . 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:

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

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. It is simulated in Ueltzhöfer and Klüppelberg (2011) who obtain a better MISE (for  $N$ ) than ours because of this singularity. Nevertheless 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 Gaussian density with variance  $\delta^2$ . Thus  $g(x) = \lambda x f(x) = \lambda x e^{-x^2/(2\delta^2)}/(\delta\sqrt{2\pi})$  and  $g^*(u) = i\lambda\delta u e^{-\delta^2 u^2/2}$ . 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 Figure 2. Note that is the jump part of the so-called Merton model used for describing the log price in financial modeling. Here we choose  $\lambda = 0.1$  and  $\delta = 0.05$ .

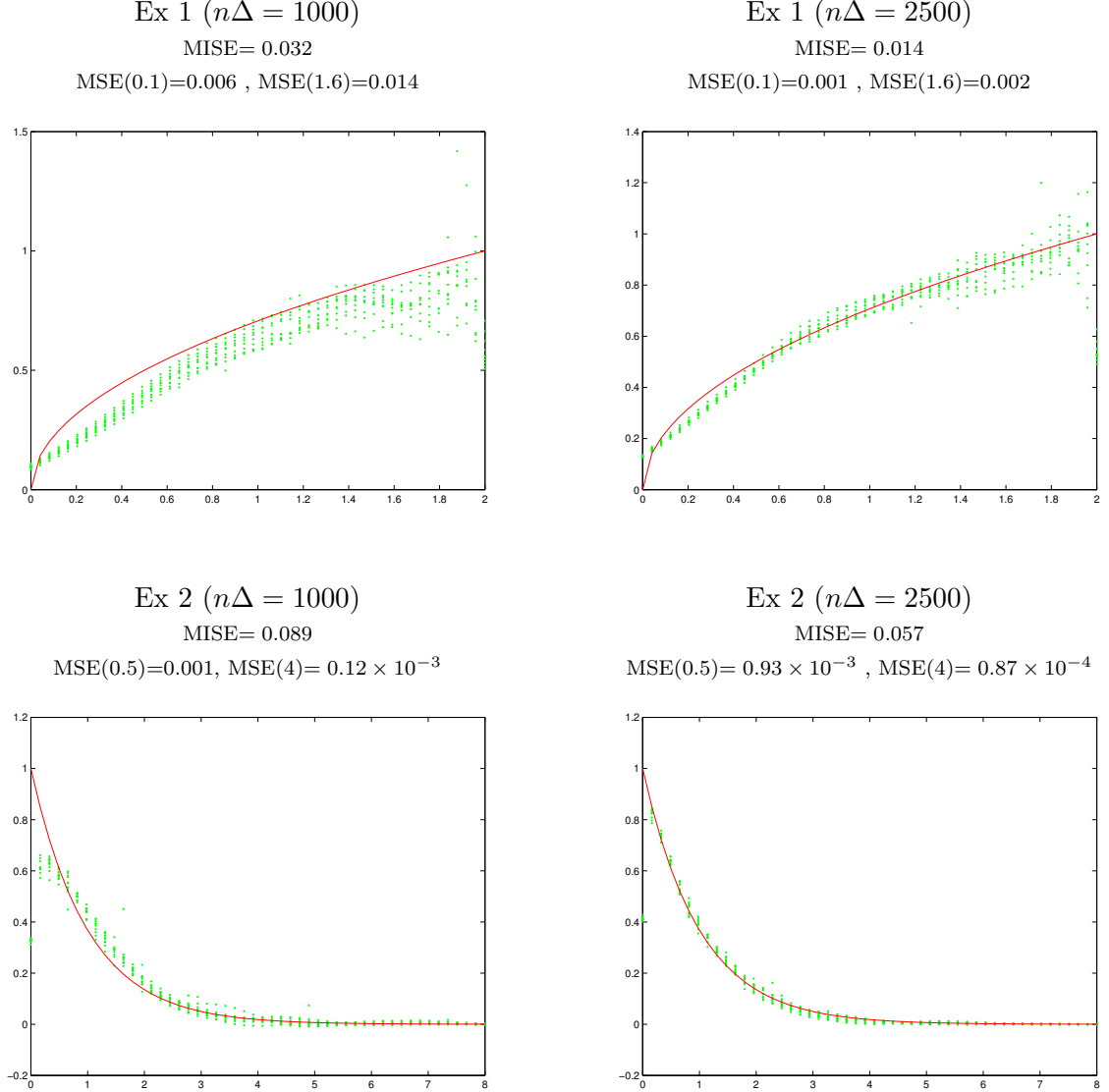
**Example 4.** Our last example is the Variance Gamma process, as described in Madan et al. (1998). It is used for modeling the dynamics of the logarithm of stock prices. The process is obtained in evaluating a Brownian motion at a time given by a Lévy-Gamma process. Denoting  $(B_t)$  a standard Brownian motion, and  $(X_t)$  a Lévy-Gamma process with parameters  $(1/\nu, 1, \nu)$  independent of  $(B_t)$ , we set  $L_t = \theta X_t + \sigma B_{X_t}$ . Then  $L_t$  is a Lévy process, with

$$g(x) = \frac{x \exp(\theta x / \sigma^2)}{\nu |x|} \exp \left( -\frac{1}{\sigma} \sqrt{\frac{2}{\nu} + \frac{\theta^2}{\sigma^2}} |x| \right).$$

As in example 2, there is a discontinuity in 0. Here we choose  $\theta = 0.1, \sigma^2 = 0.1, \nu = 0.5$ .

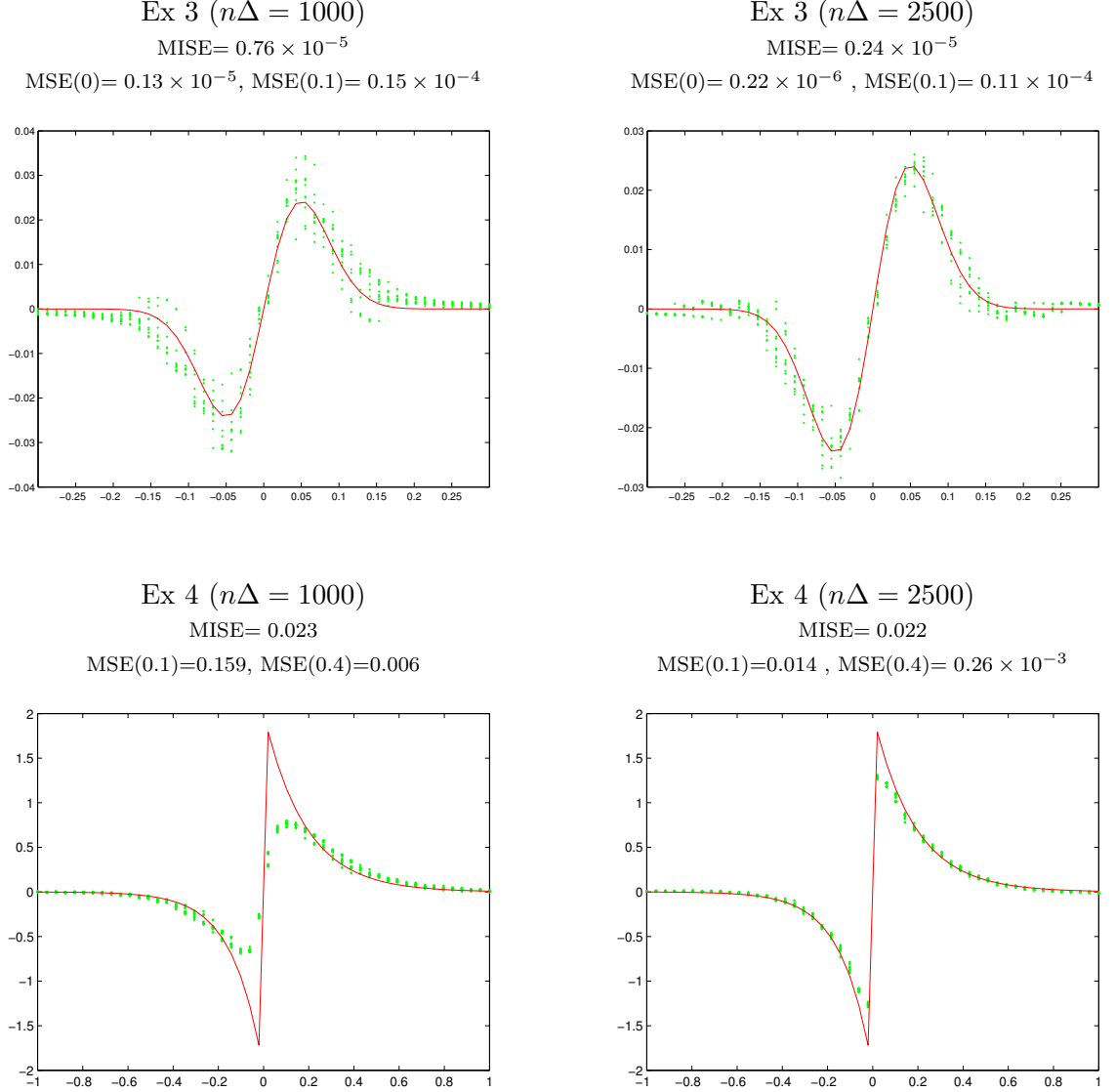
## 5. IRREGULAR SAMPLING

For high frequency data, it is frequent that the sampling is irregular, i.e. the interval  $\Delta$  is not necessarily the same at each time. In this section we consider the following framework.

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

The observations are  $(L_{t_k}, k = 1, \dots, n)$  where  $(L_t)$  is still a Lévy process with characteristic function (1). For each  $k \geq 1$ , we denote  $\Delta_k = t_k - t_{k-1}$  the sampling intervals. Notice that it includes the previous case when for each  $k$ ,  $\Delta_k = \Delta$ . The increments are denoted by  $Z_k = L_{t_k} - L_{t_{k-1}}$ . In this context of irregular sampling, they are still independent but with non-identical distribution:  $Z_k$  has the same law than  $L_{\Delta_k}$ . To define an estimator, we observe that  $\mathbb{E}[Z_k e^{iuZ_k}] = \Delta_k \psi_{\Delta_k}(u) g^*(u)$ , and then

$$\mathbb{E} \left[ \frac{1}{\sum_{k=1}^n \Delta_k} \sum_{k=1}^n Z_k e^{iuZ_k} \right] = \left( \frac{\sum_{k=1}^n \Delta_k \psi_{\Delta_k}(u)}{\sum_{k=1}^n \Delta_k} \right) g^*(u).$$

FIGURE 2. Function  $g$  (solid line) and estimators  $\hat{g}_h$  (dotted lines).

Thus, denoting  $\bar{\Delta} = \frac{1}{n} \sum_{k=1}^n \Delta_k$ , we introduce

$$(9) \quad \hat{g}_h^*(u) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k e^{iuZ_k} K^*(hu), \quad \hat{g}_h(x) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k K_h(x - Z_k).$$

Additionally, for all real  $\delta$ , we denote  $\overline{\Delta^\delta} = \frac{1}{n} \sum_{k=1}^n \Delta_k^\delta$ . We can bound the Mean Squared Error of this estimate:

**Proposition 5.1.** *Under  $G2$ ,  $G3(2)$ ,  $G4(\beta)$ ,  $G5$  and if  $K$  satisfies  $K1$  and  $K2(\beta)$ , we have*

$$(10) \quad MSE(x_0, h) \leq c_1 h^{2\beta} + c_2 \frac{1}{nh\bar{\Delta}} + c'_2 \frac{\bar{\Delta}^2}{nh\bar{\Delta}^2} + c'_1 \left( \frac{\bar{\Delta}^2}{\bar{\Delta}} \right)^2$$

with  $c_1 = 2(L/\lfloor \beta \rfloor! \int |K(v)| |v|^\beta dv)^2$ ,  $c'_1 = 2(2\|g'\|_\infty \|g\|_1 \|K\|_1)^2$ ,  $c_2 = \|(g^*)'\|_1 \|K\|_2^2 / (2\pi)$ ,  $c'_2 = \|K\|_2^2 \|g\|_2^2$ .

The proof is similar to the case of regular sampling, therefore it is omitted.

In this section, we are still interested in the high frequency context: the asymptotic framework is

**S':**  $\bar{\Delta} \rightarrow 0$  and  $n\bar{\Delta} \rightarrow \infty$  when  $n \rightarrow \infty$ . We shall also assume that  $(\bar{\Delta}^2)^2/\bar{\Delta} \leq 1$  and

$$(11) \quad \frac{(\bar{\Delta}^2)^2}{\bar{\Delta}} = O(n^{-1}).$$

Condition (11) is verified for instance if  $\Delta_k = Ck^{-\alpha}$  with  $\alpha \in [1/3, 1]$ . Then we find the same rate of convergence replacing  $\Delta$  by  $\bar{\Delta}$ :

**Proposition 5.2.** *Under the assumptions of Proposition 5.1 and under condition (11), the choice  $h_{opt} \propto ((n\bar{\Delta})^{-\frac{1}{2\beta+1}})$  minimizes the risk bound (10) and gives  $MSE(x_0, h_{opt}) = O((n\bar{\Delta})^{-\frac{2\beta}{2\beta+1}})$ .*

As already noticed in Comte and Genon-Catalot (2010a), other estimation strategies than (9) are possible. For each real  $\delta$ , we obtain an estimator by setting

$$\hat{g}_h(x) = \frac{1}{n\bar{\Delta}^{\delta+1}} \sum_{k=1}^n \Delta_k^\delta Z_k K_h(x - Z_k).$$

Under suitable conditions, this estimate has a MSE bounded by a constant times  $(n\bar{\Delta}^{\delta+1})^2 / \bar{\Delta}^{2\delta+1})^{-\frac{2\beta}{2\beta+1}}$ . But, for all  $\delta$ , by the Schwarz inequality,  $\bar{\Delta}^{\delta+1} / \bar{\Delta}^{2\delta+1} \leq \bar{\Delta}$ . That is why we prefer estimator (9).

To build an adaptive estimator, we use the same method of bandwidth selection. The set of bandwidth is still  $H = \{\frac{j}{M}, 1 \leq j \leq M\}$ . We also define

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

and we set as previously  $A(h, x_0) = \sup_{h' \in H} [|\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0)|^2 - V(h')]_+$  with

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

Then the estimator is  $\hat{g}_{\hat{h}}(x_0)$  with  $\hat{h} = \hat{h}(x_0) \in \arg \min_{h \in H} \{A(h, x_0) + V(h)\}$ .

We can state the following oracle inequality (the proof is very similar to the one of Theorem 3.2 and is therefore omitted).

**Theorem 5.1.** *We use a kernel satisfying K1 and  $M = O((n\bar{\Delta})^{1/3})$ . Assume that  $g$  satisfies G1, G2, G3(5) and take*

$$(12) \quad C_0 = \frac{c}{2\pi} \|K\|_2^2 (\|(g^*)'\|_1 + \|g^*\|_2^2)$$

with  $c \geq 16 \max(1, \|K\|_\infty)$ . Then, under scheme  $S'$ ,

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

Moreover, if  $g$  satisfies G5,  $G4(\beta)$  with  $\beta \geq 1$  and the kernel satisfying K1 and  $K2(\beta)$ , and  $M = \lfloor (n\bar{\Delta})^{1/3} \rfloor$ , then

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

Thus the rate of convergence in this case of irregular sampling is  $(\log(n\bar{\Delta})/n\bar{\Delta})^{-\frac{2\beta}{2\beta+1}}$  provided that  $(\bar{\Delta}^2)^2/\bar{\Delta} = O(n^{-1})$ .

## 6. PROOFS

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

**Proposition 6.1.** *Denote by  $P_\Delta$  the distribution of  $Z_1^\Delta$  and define  $\mu_\Delta(dx) = \Delta^{-1}xP_\Delta(dx)$ . 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 6.2.** *Let  $p \geq 1$  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,  $\mathbb{E}(|Z_1^\Delta|) \leq 2\Delta\|g\|_1$ .*

**6.1. Proof of Lemma 3.1.** First, we study  $b_2(x_0)$  using Proposition 6.1:

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

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

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

From the results of Proposition 6.2 we obtain

$$(13) \quad |b_2(x_0)| \leq 2\|g'\|_\infty \|K\|_1 \|g\|_1 \Delta.$$

To study  $b_1(x_0) = K_h \star g(x_0) - g(x_0)$ , it is sufficient to use Taylor's theorem and  $G4(\beta)$  (this is a classic computation, see Tsybakov (2009) for details) and we obtain

$$(14) \quad |b_1(x_0)| \leq \frac{h^\beta L}{l!} \int |K(v)| |v|^\beta dv.$$

Gathering (13) and (14) completes the proof of Lemma 3.1.  $\square$

**6.2. Proof of Lemma 3.2.** As the  $Z_k^\Delta$  are i.i.d., we have:

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

Thus,

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

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  $v = u/h$

$$\begin{aligned} \text{Var}[\widehat{g}(x_0)] &\leq \frac{1}{n\Delta^2} \mathbb{E} \left[ (Z_1^\Delta)^2 \left| \frac{1}{2\pi} \int K^*(vh) e^{-i(x_0 - Z_1^\Delta)v} dv \right|^2 \right] \\ &\leq \frac{1}{n\Delta^2(2\pi)^2} \mathbb{E} \left[ \iint Z_1^\Delta e^{iZ_1^\Delta v} K^*(vh) e^{-ix_0 v} \overline{Z_1^\Delta e^{iZ_1^\Delta u} K^*(uh) e^{-ix_0 u}} dv du \right]. \end{aligned}$$

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

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

Now the following formula

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

gives  $\text{Var}[\widehat{g}(x_0)] \leq T_1 + T_2$  with

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

We first bound  $T_2$ :

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

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.  $\square$

**6.3. Proof of the lower bound.** Here we prove 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

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

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  $|K^{(k)}(x)| \leq |x|^{-2-k}$  for  $|x|$  large enough (say for  $|x| > B$ ) and for all  $k \geq 0$ . We shall also use that the Fourier transform  $K^*$  exists, is differentiable almost everywhere with  $K^*$  and  $(K^*)' \in \mathbb{L}^1 \cap \mathbb{L}^2 \cap \mathbb{L}^\infty$ . Take for instance  $K$  equals to the difference between the Cauchy density and the normal density. Using this auxiliary function  $K$ , we can define

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

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

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

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

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

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

$$N_0^{-1}(x) |N_1(x) - N_0(x)| \leq C c h_n^\beta (1 + (\lambda(|x_0| + B h_n))^2) \|K\|_\infty \leq 1$$

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  $Q$ ,  $\chi^2(P, Q) = \int (dP/dQ - 1)^2 dQ$ . We shall use the following result stated in Tsybakov (2009) (section 2.2 and Theorem 2.2):

**Theorem.** *Let  $\Theta$  be a nonparametric class of functions containing the function  $\theta$  to estimate, and  $\{P_\theta, \theta \in \Theta\}$  be a family of probability measures on a measurable space  $(\mathcal{X}, \mathcal{A})$  associated with the data. Let  $d$  be a distance on  $\Theta$ . Let  $\theta_0$  and  $\theta_1$  be two functions in  $\Theta$  such that  $d(\theta, \theta_1) \geq 2\psi_n$ . If  $\chi^2(P_1, P_0) \leq \alpha < \infty$  then*

$$\inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta \left[ \psi_n^{-2} d^2(\hat{\theta}_n, \theta) \right] \geq \max \left( \frac{e^{-\alpha}}{4}, \frac{1 - \sqrt{\alpha/2}}{2} \right)$$

Then it is sufficient to 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  $Z_1^\Delta, \dots, Z_n^\Delta$  s.t the associated Lévy process  $L_0$  (resp.  $L_1$ ) has Lévy measure  $N_0(x)dx$  (resp.  $N_1(x)dx$ ).

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  $i = 0, 1$ ,  $\|g_i^{(k+1)}\|_p \leq L$  where  $k = \lfloor \beta \rfloor$  and  $p^{-1} = 1 + k - \beta$ . Indeed Hölder inequality gives

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

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  $\|g_0^{(k+1)}\|_p \leq L/2 \leq L$ .

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

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

In the same way

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

These integrals are finite since  $|K^{(k)}(v)| \leq v^{-(2+k)}$  so that  $|vK^{(k+1)}(v)| \leq v^{-(2+k)}$  for  $v$  large enough and  $p(k+2) = (k+2)/(k+1-\beta) > 1$ . Thus

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

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,  $x_0 \neq 0$  and we can see that  $|g_1(x_0) - g_0(x_0)| = ch_n^\beta |K(0)x_0|$  with  $K(0) \neq 0$ . 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

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

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

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

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

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

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

For  $i = 0, 1$ , the Fourier transform 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

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

Thus

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

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

$$(15) \quad |\psi_1(u) - \psi_0(u)| \leq e^C \Delta ch_n^{\beta+1} |K^*(h_n u)|.$$

Then

$$(16) \quad I_1 \leq C\Delta^{-1} \int \left| \Delta h_n^{\beta+1} K^*(h_n u) \right|^2 du \leq C\Delta h_n^{2\beta+1}.$$

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

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

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 except at  $u = 0$ , with derivative

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

where

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

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

$$(17) \quad \Delta^{-1} F^{*' } = \gamma_1 \psi_{1,\Delta} - \gamma_0 \psi_{0,\Delta} = (\gamma_1 - \gamma_0) \psi_{1,\Delta} + \gamma_0 (\psi_{1,\Delta} - \psi_{0,\Delta})$$

with  $|\psi_{1,\Delta}| \leq 1$ . Since  $K^*$  and  $K^{*'}$  are uniformly bounded,  $\gamma_1 - \gamma_0$  is bounded as well. In the same way, the inequality (15) entails that  $\|\psi_{1,\Delta} - \psi_{0,\Delta}\|_\infty < \infty$ , so that  $F^{*'}$  is bounded. Thus  $F^*$  is Lipschitz and absolutely continuous. Moreover, using again (17), we can see that  $F^{*'}$  is integrable. Then, according to Rudin (1987), the Fourier transform of  $F^{*'}$  is  $-ix F^{**}(x)$  (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  $(F^{*'})^*(x) = -2\pi i x F(-x)$  a.e.. Next, the Parseval equality provides  $\int |xF(x)|^2 dx = (2\pi)^{-1} \int |F^{*'}(u)|^2 du$ . Thus

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

Hence, using the factorization (17) 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

$$\begin{aligned} I_{2,1} &= c^2 h_n^{2\beta+2} \int |ix_0 K^*(h_n u) + h_n K^{*'}(h_n u)|^2 du \\ &\leq 2c^2 h_n^{2\beta+1} \left( x_0^2 \int |K^*|^2 + h_n \int |K^{*'}|^2 \right) \\ (18) \quad &\leq 4\pi c^2 h_n^{2\beta+1} \left( x_0^2 \int |K|^2 + h_n \int |xK(x)|^2 \right) \leq Ch_n^{2\beta+1}. \end{aligned}$$

Now, in order to deal with  $I_{2,2}$ , we use the previous bound (15) on  $|\psi_{1,\Delta} - \psi_{0,\Delta}|$

$$\begin{aligned} I_{2,2} &\leq Cc^2 \Delta^2 h_n^{2\beta+2} \int |\gamma_0(u) K^*(h_n u)|^2 du \\ (19) \quad &\leq Cc^2 \Delta^2 h_n^{2\beta+2} \|K^*\|_\infty^2 \|\gamma_0\|_2^2 \leq Ch_n^{2\beta+1} \end{aligned}$$

since  $\Delta$  is bounded.

Finally, by gathering (16), (18) and (19), since  $h_n = (n\Delta)^{1/(2\beta+1)}$ , 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.  $\square$

**6.4. 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

$$|g(x_0) - \hat{g}_{\hat{h}}(x_0)|^2 \leq 3R_1^2 + 3R_2^2 + 3R_3^2$$

with  $R_1 = |\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)|$ ,  $R_2 = |\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)|$ ,  $R_3 = |\hat{g}_h(x_0) - g(x_0)|$ . According to the definition of  $A(h)$ :

$$A(h) \geq |\hat{g}_{\hat{h}}(x_0) - \hat{g}_{h,\hat{h}}(x_0)|^2 - V(\hat{h}) = R_1^2 - V(\hat{h}).$$

So  $R_1^2 \leq A(h) + V(\hat{h})$ . In the same way,  $A(\hat{h}) \geq |\hat{g}_{h,\hat{h}}(x_0) - \hat{g}_h(x_0)|^2 - V(h) = R_2^2 - V(h)$ .

So  $R_2^2 \leq A(\hat{h}) + V(h)$ . 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  $\hat{h}$ ,  $A(\hat{h}) + V(\hat{h}) \leq A(h) + V(h)$ . This allows us to write

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

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 notation as in Lemma 3.1, but with subscript  $h$ ). Thus

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

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

$$(20) \quad \hat{g}_{h,h'} - \hat{g}_{h'} = \hat{g}_{h,h'} - g_{h,h'} - \hat{g}_{h'} + g_{h'} + g_{h,h'} - g_{h'},$$

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:

$$g_{h,h'}(x_0) - g_{h'}(x_0) = K_{h'} \star b_h(x_0) - b_{h',2}(x_0).$$

Now, using inequality (13),  $|b_{h',2}(x_0)| \leq 2\|g'\|_\infty \|K\|_1 \|g\|_1 \Delta$ , so that

$$\begin{aligned} |g_{h,h'}(x_0) - g_{h'}(x_0)|^2 &\leq 2|K_{h'} \star b_h(x_0)|^2 + O(\Delta^2) \\ (21) \quad &\leq 2\|K\|_1^2 \text{ess sup } |b_h|^2 + O(\Delta^2). \end{aligned}$$

Then by inserting (21) in decomposition (20), we find:

$$\begin{aligned}
 A(h) &= \sup_{h'} \{|\hat{g}_{h,h'}(x_0) - \hat{g}_{h'}(x_0)|^2 - V(h')\}_+ \\
 &\leq 3 \sup_{h'} \{|\hat{g}_{h,h'}(x_0) - g_{h,h'}(x_0)|^2 - V(h')/6\}_+ \\
 (22) \quad &+ 3 \sup_{h'} \{|\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - V(h')/6\}_+ + 6\|K\|_1^2 \text{ess sup } |b_h|^2 + O(\Delta^2).
 \end{aligned}$$

We can prove the following concentration result:

**Proposition 6.3.** *Assume that  $g$  satisfies  $G1$ ,  $G2$ ,  $G3(5)$ ,  $K$  satisfies  $K1$ ,  $M = O((n\Delta)^{1/3})$ ,  $\Delta \leq 1$  and take  $c$  in (7) such that  $c \geq 16 \max(1, \|K\|_\infty)$ . Then*

$$(23) \quad \mathbb{E} \left[ \sup_{h'} \{|\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - V(h')/6\}_+ \right] = O \left( \frac{\log(n\Delta)}{n\Delta} \right),$$

$$(24) \quad \mathbb{E} \left[ \sup_{h'} \{|\hat{g}_{h,h'}(x_0) - g_{h,h'}(x_0)|^2 - V(h')/6\}_+ \right] = O \left( \frac{\log(n\Delta)}{n\Delta} \right).$$

This proposition is proved in Section 6.6 page 23. Inequalities (23) et (24) together with (22) imply

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2] \leq C_1 \text{ess sup } |b_h|^2 + C_2 V(h) + C_3 \frac{\log(n\Delta)}{n\Delta}.$$

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

**6.5. Proof of Theorem 3.3.** In all this proof, we shall use the following notation:

$$\hat{\theta}_\Delta(u) = \frac{1}{n} \sum_{k=1}^n Z_k^\Delta e^{iZ_k^\Delta u}, \quad \hat{\eta}_\Delta(u) = \frac{1}{n} \sum_{k=1}^n (Z_k^\Delta)^2 e^{iZ_k^\Delta u},$$

and  $\theta_\Delta(u) = \mathbb{E}\hat{\theta}_\Delta(u)$ ,  $\eta_\Delta(u) = \mathbb{E}\hat{\eta}_\Delta(u)$ . We also denote  $f(x) = xg(x)$ , so that  $f^*(u) = -i(g^*)'(u)$  is estimated by  $\hat{f}_{h_1}^* = \hat{\eta}_\Delta(u)K^*(uh_1)/\Delta$ . We shall use the following Lemma.

**Lemma 6.1** (Proposition 2.3 in Comte and Genon-Catalot (2009)). *Assume that  $g$  is integrable, then we have:*

$$(25) \quad |\psi_\Delta(u) - 1| \leq |u|\Delta\|g\|_1.$$

Moreover under  $G3(2p)$ , for  $p \geq 1$ ,

$$(26) \quad \Delta^{-2p} \mathbb{E} \left| \hat{\theta}_\Delta(v) - \theta_\Delta(v) \right|^{2p} \leq C(n\Delta)^{-p}.$$

Now, let

$$\Omega = \{\|g^* - \hat{g}_{h_2}^*\|_2 \leq \|g^*\|_2(1 - 1/\sqrt{2}) \quad \text{and} \quad \|f^* - \hat{f}_{h_1}^*\|_1 \leq \|f^*\|_1/2\},$$

where  $h_1 = (n\Delta)^{-1/3} = h_2$ , as defined page 7. The proof is decomposed in three steps. First we shall prove that the inequality is true on  $\Omega$ , i.e.

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2 \mathbf{1}_\Omega] \leq C \left\{ \inf_{h \in H} \left\{ \text{ess sup } |g - \mathbb{E}[\hat{g}_h]|^2 + V(h) \right\} + \frac{\log(n\Delta)}{n\Delta} \right\}.$$

The second step is to show the rough upper bound

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^4] \leq C(n\Delta)^{2/3}.$$

Finally we will show that  $\mathbb{P}(\Omega^c) \leq C(n\Delta)^{-8/3}$ . Consequently

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2 \mathbf{1}_{\Omega^c}] \leq \sqrt{\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^4] \mathbb{P}(\Omega^c)} \leq C(n\Delta)^{-1}$$

and the theorem is proved.

• *First step:*

Following the proof of Theorem 3.2, we can obtain

$$|g(x_0) - \hat{g}_h(x_0)|^2 \mathbf{1}_\Omega \leq 6A(h) \mathbf{1}_\Omega + 6\hat{V}(h) \mathbf{1}_\Omega + 3|g(x_0) - \hat{g}_h(x_0)|^2.$$

Now, let us remark that on  $\Omega$

$$\frac{1}{2} \|g^*\|_2^2 \leq \|\hat{g}_{h_2}^*\|_2^2 \leq (2 - 1/\sqrt{2})^2 \|g^*\|_2^2 \quad \text{and} \quad \frac{1}{2} \|f^*\|_1 \leq \|\hat{f}_{h_1}^*\|_1 \leq \frac{3}{2} \|f^*\|_1$$

with  $\|f^*\|_1 = \|(g^*)'\|_1$ , so that on  $\Omega$ ,  $\frac{1}{2}C_0 \leq \widehat{C}_0 \leq 2C_0$  and

$$(27) \quad \frac{1}{2} V(h) \mathbf{1}_\Omega \leq \hat{V}(h) \mathbf{1}_\Omega \leq 2V(h)$$

We thus get

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2 \mathbf{1}_\Omega] \leq 6\mathbb{E}[A(h) \mathbf{1}_\Omega] + 12V(h) + 3\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2],$$

which leads to

$$\mathbb{E}[|g(x_0) - \hat{g}_h(x_0)|^2 \mathbf{1}_\Omega] \leq 6\mathbb{E}[A(h) \mathbf{1}_\Omega] + 3b_h^2(x_0) + (12 + 3/c) V(h).$$

Using the definition of  $A(h)$ , it is then sufficient to prove

$$(28) \quad \mathbb{E} \left[ \sup_{h'} \{ |\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - \hat{V}(h')/6 \}_+ \mathbf{1}_\Omega \right] = O \left( \frac{\log(n\Delta)}{n\Delta} \right),$$

$$(29) \quad \mathbb{E} \left[ \sup_{h'} \{ |\hat{g}_{h,h'}(x_0) - g_{h,h'}(x_0)|^2 - \hat{V}(h')/6 \}_+ \mathbf{1}_\Omega \right] = O \left( \frac{\log(n\Delta)}{n\Delta} \right)$$

to obtain the result. Using (27) and Proposition 6.3, since  $c/2 \geq 16 \max(1, \|K\|_\infty)$ ,

$$\begin{aligned} & \mathbb{E} \left[ \sup_{h'} \{ |\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - \hat{V}(h')/6 \}_+ \mathbf{1}_\Omega \right] \\ & \leq \mathbb{E} \left[ \sup_{h'} \left\{ |\hat{g}_{h'}(x_0) - g_{h'}(x_0)|^2 - \frac{1}{6} \frac{c/2}{2\pi} \|K\|^2 (\|(g^*)'\|_1 + \|g^*\|_2^2) \frac{\log(n\Delta)}{n\Delta} \right\}_+ \right] \\ & = O \left( \frac{\log(n\Delta)}{n\Delta} \right) \end{aligned}$$

and we prove (29) in the same way.

• *Second step:*

First, using Lemma 3.1,  $|g_h(x_0) - g(x_0)|^2 \leq \sup_{h \in H} (c_1 h^2 + c'_1 \Delta^2) \leq C$ . Then the bias term is uniformly bounded. Let us now study the “variance” term. We can write

$$\hat{g}_h(x_0) = \frac{1}{2\pi} \int e^{-ix_0 u} K^*(uh) \frac{1}{\Delta} \hat{\theta}_\Delta(u) du$$

and, since all  $h \in H$  is larger than  $1/M$ ,

$$\begin{aligned} |\hat{g}_h(x_0) - g_h(x_0)| &\leq \frac{1}{2\pi} \sup_{h \in H} \int |K^*(uh)| \left| \frac{\hat{\theta}_\Delta(u) - \theta_\Delta(u)}{\Delta} \right| du \\ &\leq \frac{M}{2\pi} \sum_{h \in H} \int |K^*(u)| \left| \frac{\hat{\theta}_\Delta(u/h) - \theta_\Delta(u/h)}{\Delta} \right| du. \end{aligned}$$

With a convex inequality

$$|\hat{g}_h(x_0) - g_h(x_0)|^4 \leq \frac{M^7}{(2\pi)^4} \sum_{h \in H} \left( \int |K^*(u)| \left| \frac{\hat{\theta}_\Delta(u/h) - \theta_\Delta(u/h)}{\Delta} \right| du \right)^4$$

Next, we use the following inequality (obtained with two uses of the Schwarz inequality):

$$\begin{aligned} \mathbb{E} \left[ \left( \int \phi(u) du \right)^4 \right] &= \iiint \mathbb{E} [\phi(u_1) \dots \phi(u_4)] du_1 \dots du_4 \\ &\leq \iiint \mathbb{E}^{1/4} [\phi(u_1)^4] \dots \mathbb{E}^{1/4} [\phi(u_4)^4] du_1 \dots du_4 = \left( \int \mathbb{E}^{1/4} [\phi(u)^4] du \right)^4 \end{aligned}$$

with  $\phi(u) = |K^*(u)| \left| \frac{\hat{\theta}_\Delta(u/h) - \theta_\Delta(u/h)}{\Delta} \right|$ . Thus,

$$\mathbb{E} [|\hat{g}_h(x_0) - g_h(x_0)|^4] \leq \frac{M^7}{(2\pi)^4} \sum_{h \in H} \left( \int |K^*(u)| \mathbb{E}^{1/4} \left[ \left| \frac{\hat{\theta}_\Delta(u/h) - \theta_\Delta(u/h)}{\Delta} \right|^4 \right] du \right)^4.$$

But, according to (26) in Lemma 6.1, under  $G3(4)$ ,  $\Delta^{-4} \mathbb{E} \left| \hat{\theta}_\Delta(v) - \theta_\Delta(v) \right|^4 \leq C(n\Delta)^{-2}$ . Hence, under  $G3(4)$ ,

$$\begin{aligned} \mathbb{E} |\hat{g}_h(x_0) - g_h(x_0)|^4 &\leq CM^7 \sum_{h \in H} \left( \int |K^*(u)| (n\Delta)^{-1/2} du \right)^4 \\ &\leq C \|K^*\|_1^4 M^8 (n\Delta)^{-2} \leq C \|K^*\|_1^4 (n\Delta)^{2/3}. \end{aligned}$$

• *Third step:*

$$\begin{aligned} \mathbb{P}(\Omega^c) &= \mathbb{P}(\|g^* - \hat{g}_{h_2}^*\|_2 > \|g^*\|_2(1 - 1/\sqrt{2}) \text{ or } \|f^* - \hat{f}_{h_1}^*\|_1 > \|f^*\|_1/2) \\ &\leq (\|g^*\|_2(1 - 1/\sqrt{2}))^{-8} \mathbb{E} \|\hat{g}_{h_2}^* - g^*\|_2^8 + (\|f^*\|_1/2)^{-16} \mathbb{E} \|\hat{f}_{h_1}^* - f^*\|_1^{16} \\ &\leq C \left( \mathbb{E} \|\hat{g}_{h_2}^* - g_{h_2}^*\|_2^8 + \mathbb{E} \|g_{h_2}^* - g^*\|_2^8 + \mathbb{E} \|\hat{f}_{h_1}^* - f_{h_1}^*\|_1^{16} + \mathbb{E} \|f_{h_1}^* - f^*\|_1^{16} \right). \end{aligned}$$

Thus we have four terms to upperbound.

**First term:** Since  $\hat{g}_{h_2}^*(u) = K_0^*(uh_2)\hat{\theta}_\Delta(u)/\Delta$ ,

$$\|\hat{g}_{h_2}^* - g_{h_2}^*\|_2^2 = \frac{1}{h_2} \int |K_0^*(u)|^2 \left| \frac{\hat{\theta}_\Delta(u/h_2) - \theta_\Delta(u/h_2)}{\Delta} \right|^2 du.$$

Then, according to (26) in Lemma 6.1, under  $G3(8)$ ,

$$\begin{aligned} \mathbb{E}\|\hat{g}_{h_2}^* - g_{h_2}^*\|_2^8 &\leq \frac{1}{h_2^4} \left( \int \mathbb{E}^{1/4} \left[ |K_0^*(u)|^8 \left| \frac{\hat{\theta}_\Delta(u/h_2) - \theta_\Delta(u/h_2)}{\Delta} \right|^8 \right] du \right)^4 \\ &\leq \frac{1}{h_2^4} \left( \int |K_0^*(u)|^2 (n\Delta)^{-1} du \right)^4 \leq \|K_0^*\|_2^8 M^4 (n\Delta)^{-4} \leq 16(n\Delta)^{-8/3}. \end{aligned}$$

**Second term:** Since  $g_{h_2}^* = K_0^*(uh_2)g^*(u)\psi_\Delta(u)$ , we can decompose the bias into

$$g^*(u) - g_{h_2}^*(u) = g^*(u)(1 - K_0^*(uh_2)) + g^*(u)K_0^*(uh_2)(1 - \psi_\Delta(u)) = b_1 + b_2.$$

Using that  $g \in \text{Sob}(1)$  ( $\int |g^*(u)|^2 u^2 du < \infty$ ),

$$\begin{aligned} \|b_1\|_2^2 &= \int |g^*(u)(1 - K_0^*(uh_2))|^2 du = \int |g^*(u)|^2 \mathbb{1}_{|uh_2|>1} du \\ &\leq \int |g^*(u)|^2 |uh_2|^2 du \leq Ch_2^2. \end{aligned}$$

On the other hand, using (25) in Lemma 6.1,

$$\begin{aligned} \|b_2\|_2^2 &= \int |g^*(u)K_0^*(uh_2)(1 - \psi_\Delta(u))|^2 du \leq C\Delta^2 \int |g^*(u)u|^2 du \\ &\leq C\Delta^2 \leq C(n\Delta)^{-1}. \end{aligned}$$

Thus, taking  $h_2 = (n\Delta)^{-1/3}$  gives  $\|g^* - g_{h_2}^*\|^8 \leq Ch_2^8 + C(n\Delta)^{-4} \leq C(n\Delta)^{-8/3}$ .

**Third term:** Since  $\hat{f}_{h_1}^*(u) = K_0^*(uh_1)\hat{\eta}_\Delta(u)/\Delta$ ,

$$\|\hat{f}_{h_1}^* - f_{h_1}^*\|_1 \leq \frac{1}{h_1} \int |K_0^*(u)| \left| \frac{\hat{\eta}_\Delta(u/h_1) - \eta_\Delta(u/h_1)}{\Delta} \right| du.$$

Next, we use the following inequality

$$\mathbb{E} \left[ \left( \int \phi(u) du \right)^{16} \right] \leq \left( \int \mathbb{E}^{1/16} [\phi(u)^{16}] du \right)^{16}.$$

Exactly as Lemma 6.1, using the Rosenthal inequality, we can prove under  $G3(4p)$ , for  $p \geq 1$ ,  $\Delta^{-2p} \mathbb{E} |\hat{\eta}_\Delta(v) - \eta_\Delta(v)|^{2p} \leq C(n\Delta)^{-p}$ . Then, under  $G3(32)$ ,

$$\begin{aligned} \mathbb{E}\|\hat{f}_{h_1}^* - f_{h_1}^*\|_1^{16} &\leq \frac{1}{h_1^{16}} \left( \int \mathbb{E}^{1/16} \left[ |K_0^*(u)|^{16} \left| \frac{\hat{\eta}_\Delta(u/h_1) - \eta_\Delta(u/h_1)}{\Delta} \right|^{16} \right] du \right)^{16} \\ &\leq \frac{1}{h_1^{16}} \left( \int |K_0^*(u)| (n\Delta)^{-1/2} du \right)^{16} \leq C\|K^*\|_1 (n\Delta)^{-8/3} \end{aligned}$$

since  $h_1 = (n\Delta)^{-1/3}$ .

**Fourth term:** Since  $\eta_\Delta = -\psi''_\Delta = \Delta f^* \psi_\Delta + \Delta^2 (g^*)^2 \psi_\Delta$ , we can decompose the bias into

$$\begin{aligned} f^*(u) - f_{h_1}^*(u) &= f^*(u) - K_0^*(uh_1)f^*(u)\psi_\Delta(u) - \Delta K_0^*(uh_1)(g^*(u))^2\psi_\Delta(u) \\ &= f^*(u)(1 - K_0^*(uh_1)) + f^*(u)K_0^*(uh_1)(1 - \psi_\Delta(u)) \\ &\quad - \Delta K_0^*(uh_1)(g^*(u))^2\psi_\Delta(u) \\ &= b_1 + b_2 + b_3. \end{aligned}$$

Since  $xg(x) \in \text{Sob}(1)$  ( $\int |f^*(u)|^2 u^2 du < \infty$ ),

$$\begin{aligned} \|b_1\|_1 &\leq \int |f^*(u)(1 - K_0^*(uh_1))| du = \int |f^*(u)| \mathbf{1}_{|uh_1|>1} du \\ &\leq \left( \int |f^*(u)|^2 |uh_1|^2 du \int |uh_1|^{-2} \mathbf{1}_{|uh_1|>1} du \right)^{1/2} \leq Ch_1^{1/2}. \end{aligned}$$

On the other hand, using (25) in Lemma 6.1,

$$\begin{aligned} \|b_2\|_1 &\leq \int |f^*(u)K_0^*(uh_1)(1 - \psi_\Delta(u))| du \leq C\Delta \int |f^*(u)uK_0^*(uh_1)| du \\ &\leq C\Delta \left( \int |f^*(u)u|^2 du \int |K_0^*(uh_1)|^2 du \right)^{1/2} \\ &\leq C\Delta h_1^{-1/2} \leq C(h_1 n \Delta)^{-1/2}, \end{aligned}$$

and

$$\begin{aligned} \|b_3\|_1 &\leq \Delta \int |K_0^*(uh_1)(g^*(u))^2\psi_\Delta(u)| du \\ &\leq \Delta \int |(g^*(u))^2| du \leq C\Delta \leq C(n\Delta)^{-1/2}. \end{aligned}$$

Thus  $\|f^* - f_{h_1}^*\|_1^{16} \leq Ch_1^8 + C(h_1 n \Delta)^{-8} + C(n\Delta)^{-8} \leq C(n\Delta)^{-8/3}$ .

This completes the proof of Theorem 3.3.  $\square$

**6.6. Proof of Proposition 6.3.** Note that

$$(30) \quad \hat{g}_{h'}(x_0) - g_{h'}(x_0) = \frac{1}{n} \sum_{k=1}^n \left[ \frac{Z_k^\Delta}{\Delta} K_{h'}(x_0 - Z_k^\Delta) - \mathbb{E} \left( \frac{Z_k^\Delta}{\Delta} K_{h'}(x_0 - Z_k^\Delta) \right) \right].$$

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:

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

where

$$(31) \quad \mu_n = \mu_n(h') = \frac{\|K\|_2^2 (\|(g^*)'\|_1 + \|g^*\|_2^2)}{2\pi \|K\|_\infty \sqrt{V(h')/6}}.$$



We then decompose (30) as follows

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

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

$$(32) \quad W_k(h) = \frac{Z_k^\Delta}{\Delta} K_h(x_0 - Z_k^\Delta) \mathbf{1}_{\{|Z_k^\Delta| \leq \mu_n(h)\}},$$

$$(33) \quad T_k(h) = \frac{Z_k^\Delta}{\Delta} K_h(x_0 - Z_k^\Delta) \mathbf{1}_{\{|Z_k^\Delta| > \mu_n(h)\}}.$$

Thus

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

Then we use the two following lemmas

**Lemma 6.2.** *Assume that  $g$  satisfies G1, G2,  $K$  satisfies K1, and  $c \geq 16$ ,  $M = O((n\Delta)^{1/3})$ ,  $\Delta \leq 1$ . Then there exists  $C > 0$  only depending on  $K$  and  $g$  such that*

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

**Lemma 6.3.** *Under assumptions K1, G3(5) and if  $M = O((n\Delta)^{1/3})$ ,*

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

Lemmas 6.2 and 6.3 yield

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

Inequality (24) is obtained by following the same lines as for inequality (23) with  $K_h$  replaced by  $K_{h'} \star K_h$ . This ends the proof of Proposition 6.3.  $\square$

**6.7. Proof of lemma 6.2.** First, note that

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

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

**Lemma 6.4.** *Let  $W_1, \dots, W_n$   $n$  independent and identically distributed random variables and  $S_n(W) = (1/n) \sum_{i=1}^n [W_i - \mathbb{E}(W_i)]$ . Then, for  $\eta > 0$ ,*

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

where  $\text{Var}(W_1) \leq \nu^2$  and  $|W_1| \leq b$ .

We apply this form of Bernstein inequality to  $W_i(h)$  defined by (32) and  $\eta = \sqrt{(1/12 + y)V(h)}$ . Using Lemma 3.2 and  $\Delta \leq 1$ , it is easy to see that

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

We find

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

and

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

Then we deduce

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

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

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

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

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

as soon as  $c \geq 16$ . This completes the proof of lemma 6.2.  $\square$

**6.8. Proof of lemma 6.3.** For a fixed bandwidth  $h$  in  $H$ , we can establish the following bound:

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

for any  $w > 0$ . Recall that, according to Proposition 6.2,  $\mathbb{E}[|Z_1^\Delta|^{w+2}/\Delta]$  is bounded under  $G3(w+2)$ . We search conditions for  $\sum_h h^{-2} \mu_n^{-w} \leq \text{constant}$ . 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 6.3.  $\square$

#### ACKNOWLEDGEMENT

The authors thank Fabienne Comte and Valentine Genon-Catalot 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. (1998). Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375.
- Bücher, A. and Vetter, M. (2013). Nonparametric inference on Lévy measures and copulas. *Ann. Statist.*, 41(3):1485–1515.
- 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. (2010a). Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Stat. Neerl.*, 64(3):290–313.
- Comte, F. and Genon-Catalot, V. (2010b). 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. (2009a). Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*, volume 57 of *IMS Lecture Notes Monogr. Ser.*, pages 117–146. Inst. Math. Statist., Beachwood, OH.
- Figueroa-López, J. E. (2009b). Nonparametric estimation of time-changed lévy models under high-frequency data. *Adv. Appl. Probab.*, 41(4):1161–1188.
- 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.
- Goldenshluger, A. and Lepski, O. (2013). General selection rule from a family of linear estimators. *Theory of Probability & Its Applications*, 57(2):209–226.
- Gugushvili, S. (2012). Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(1):282–307.
- Jongbloed, G. and van der Meulen, F. H. (2006). Parametric estimation for subordinators and induced OU processes. *Scand. J. Statist.*, 33(4):825–847.
- Kappus, J. (2012). Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes. *SFB 649 discussion paper, No. 2012-016*.
- Kerkycharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields*, 121:137–170.
- Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The variance gamma process and option pricing. *Eur. Finance Rev.*, 2(1):79–105.
- Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248.
- Nickl, R. and Reiß, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.*, 263(10):3306–3332.
- 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.
- Ueltzhöfer, F. A. and Klüppelberg, C. (2011). An oracle inequality for penalised projection estimation of lévy densities from high-frequency observations. *Journal of Nonparametric Statistics*, 23(4):967–989.
- 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.