

Adaptive Estimation for Lévy processes

Fabienne Comte, Valentine Genon-Catalot

► **To cite this version:**

Fabienne Comte, Valentine Genon-Catalot. Adaptive Estimation for Lévy processes. Lévy Matters IV, IV, Springer International Publishing, pp.77-177, 2015, Estimation for Discretely Observed Lévy Processes, 978-3-319-12372-1. 10.1007/978-3-319-12373-8 . hal-00913367

HAL Id: hal-00913367

<https://hal.archives-ouvertes.fr/hal-00913367>

Submitted on 3 Dec 2013

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.

Copyright

Adaptive Estimation for Lévy processes.

Fabienne Comte and Valentine Genon-Catalot

Abstract This chapter is concerned with nonparametric estimation of the Lévy density of a Lévy process. The sample path is observed at n equispaced instants with sampling interval Δ . We develop several nonparametric adaptive methods of estimation based on deconvolution, projection and kernel. The asymptotic framework is: n tends to infinity, $\Delta = \Delta_n$ tends to 0 while $n\Delta_n$ tends to infinity (high frequency). Bounds for the \mathbb{L}^2 -risk of estimators are given. Rates of convergence are discussed. Estimation of the drift and Gaussian component coefficients is studied. A specific method for the estimating the jump density of compound Poisson processes is presented. Examples and simulation results illustrate the performance of estimators.

AMS Subject Classification 2000:

Primary: 62G05-62M05.

Secondary: 60G51

Key words: adaptive estimation, deconvolution, Lévy process, Non parametric projection estimators, High frequency data, kernel estimators, compound Poisson process

Fabienne Comte

MAP5, UMR CNRS 8145, University Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints-Pères, 75006 Paris, e-mail: fabienne.comte@parisdescartes.fr; homepage: <http://http://www.math-info.univ-paris5.fr/~comte/>

Valentine Genon-Catalot

MAP5, UMR CNRS 8145, University Paris Descartes, Sorbonne Paris Cité, 45 rue des Saints-Pères, 75006 Paris e-mail: valentine.genon-catalot@parisdescartes.fr; homepage: <http://http://www.math-info.univ-paris5.fr/~genon/>

1 Introduction

The aim of this chapter is to present statistical adaptive methods of estimation of the Lévy measure of a Lévy process, *i.e.* a continuous time process with stationary independent increments whose sample paths are right-continuous with left-hand limits. We refer to [9] or [58] for a detailed probabilistic study of these processes. In what follows, we assume that the process is real-valued, discretely observed at equispaced instants and inference is based on a sample of n observations.

The distribution of a Lévy process is usually specified by its characteristic triple, the drift, the Gaussian component and the Lévy measure rather than by the distribution of its independent increments. Indeed, the distributions of increments often have no closed form formula. This is why statistical references have increasingly focused on nonparametric methods. In here, we especially develop nonparametric adaptive methods and rely mainly on the papers [17], [18], [19], [20].

In statistical inference for discretely observed continuous time processes, it is now classical to distinguish two points of view. In the low frequency point of view, the sampling interval is kept fixed and asymptotic results are given as n tends to infinity. In the high frequency (HF) point of view, which is our concern here, the sampling interval tends to 0 and the total length time where observations are taken tends to infinity. The HF point of view is simpler and allows to apply to Lévy processes several adaptive methods of estimation: deconvolution, projection or kernel methods.

Section 2 gives notations and preliminary assumptions. In Section 3, moment and small sample properties are stated. Section 4 deals with pure jump Lévy processes with finite variation on compact sets and no drift. Section 5 concerns the case of Lévy processes with no Gaussian component and Section 6 the general case. In Section 7, the estimation of the drift and Gaussian component coefficients is studied. Examples are given in Section 8. Estimation procedures are illustrated on simulated data in Section 9. In Section 10, we describe a specific method for the special case of compound Poisson processes. Section 11 is devoted to bibliographic comments.

2 Notations and preliminary assumptions

Let us introduce some notations and assumptions which are successively considered. The Lévy process is denoted by (L_t) and the observations are $(L_{k\Delta}, k = 1, \dots, n)$ where Δ is the sampling interval. The statistical procedure is based on the *i.i.d.* increments $Z_k^\Delta = L_{k\Delta} - L_{(k-1)\Delta}$. We assume that, as n tends to infinity,

$$\Delta = \Delta_n \rightarrow 0, \quad \text{and} \quad n\Delta_n \rightarrow +\infty. \quad (2.1)$$

For simplicity, we omit the dependence on n and set $Z_k^\Delta = Z_k$. We assume that the Lévy measure admits a density denoted by $n(\cdot)$. The characteristic function of L_t is denoted by

$$\varphi_t(u) = \exp t \psi(u)$$

where the characteristic exponent is given by

$$\psi(u) = iu\tilde{b} - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iux1_{|x|\leq 1}) n(x)dx, \quad (2.2)$$

with $\tilde{b} \in \mathbb{R}$, $\sigma^2 \geq 0$. The Lévy density satisfies the usual assumption:

$$\int_{\mathbb{R}} (x^2 \wedge 1) n(x)dx < +\infty. \quad (2.3)$$

Thus, $(Z_k, k = 1, \dots, n)$ is an *i.i.d.* sample with characteristic function φ_{Δ} . The non-parametric estimation of $n(\cdot)$ and the estimation of the other parameters \tilde{b}, σ^2 are investigated under different sets of assumptions on the Lévy process. Depending on the assumptions, we consider the estimation of the following functions:

$$g(x) = x n(x), \quad \ell(x) = x^2 n(x), \quad p(x) = x^3 n(x). \quad (2.4)$$

2.1 Pure jump case

We first study the estimation of g , $g(x) = xn(x)$, (hence of ℓ, p) under the assumption:

$$(H1-g) \quad \int_{\mathbb{R}} |x|n(x)dx < \infty, \quad \tilde{b} = \int_{|x|\leq 1} x n(x)dx, \quad \sigma^2 = 0.$$

When the Lévy process is self-decomposable, the function g is called the canonical function and is decreasing (see [3] and [43]). Under (H1-g), the process (L_t) has finite variation on compact sets, is of pure jump type, with no drift component. Formula (2.2) simplifies into

$$\psi(u) = \int_{\mathbb{R}} (e^{iux} - 1) n(x)dx, \quad (2.5)$$

The distribution of (L_t) is therefore completely specified by the knowledge of $n(\cdot)$ which describes the jumps behavior. The process (L_t) can be written as

$$L_t = \int_{]0,t]} \int_{\mathbb{R}/\{0\}} x \hat{p}(du, dx) = \sum_{s \leq t} \Delta L_s, \quad \text{where } \Delta L_s = L_s - L_{s-}, \quad (2.6)$$

where $\hat{p}(du, dx) = \sum_{s \geq 0} \mathbb{1}_{\Delta L_s \neq 0} \delta_{s, \Delta L_s}(du, dx)$ is the random Poisson measure associated with the jumps of (L_t) with intensity $du n(x)dx$. Note that (2.6) holds under the assumption $\int_{\mathbb{R}} (|x| \wedge 1) n(x)dx < \infty$. Assumption (H1-g) is stronger and ensures that $\mathbb{E}(|L_t|) < +\infty$ with

$$\mathbb{E}(L_t) = t \int_{\mathbb{R}} x n(x)dx.$$

2.2 Case of no Gaussian component

Then, we study the estimation of ℓ , $\ell(x) = x^2 n(x)$, (hence of p and g except near the origin) under the assumption:

$$(H1-\ell) \quad \int_{\mathbb{R}} x^2 n(x) dx < \infty, \quad \sigma^2 = 0.$$

The first part of this assumption, stronger than (2.3) was proposed by [56] and is useful for statistical inference. First, for all t , $\mathbb{E}L_t^2 < +\infty$. Second, $\int_{\mathbb{R}} (e^{iux} - 1 - iux)n(x)dx$ is well defined, consequently the following expression for (2.2) holds:

$$\psi(u) = iub + \int_{\mathbb{R}} (e^{iux} - 1 - iux)n(x)dx, \quad (2.7)$$

where $b = \tilde{b} + \int_{|x|>1} xn(x)dx = \mathbb{E}L_1$ has a statistical meaning (contrary to \tilde{b}). Thus, the sample path can be expressed as:

$$L_t = bt + X_t, \quad (2.8)$$

where (X_t) is a centered square integrable pure-jump martingale:

$$X_t = \int_{]0,t]} \int_{\mathbb{R}/\{0\}} x \hat{p}(du, dx) - du n(x)dx,$$

and $\hat{p}(du, dx)$ is the random Poisson measure associated with the jumps of (L_t) (or (X_t)).

2.3 General case

Finally, we study the estimation of p , $p(x) = x^3 n(x)$, (hence of g, ℓ except near the origin) under the assumption:

$$(H1-p) \quad \int_{\mathbb{R}} |x|^3 n(x) dx < \infty.$$

Here, $\mathbb{E}|L_t|^3 < +\infty$,

$$\psi(u) = iub - \frac{1}{2} \sigma^2 u^2 + \int_{\mathbb{R}} (e^{iux} - 1 - iux)n(x)dx, \quad (2.9)$$

and

$$L_t = bt + \sigma W_t + X_t, \quad (2.10)$$

with (X_t) as above and (W_t) is a Wiener process independent of (X_t) . The estimation of b in the second case (resp. (b, σ^2) in the third case) is detailed in Section 7.

The following notations are used below. For $u : \mathbb{R} \rightarrow \mathbb{C}$ integrable, we denote its \mathbb{L}^1 -norm and its Fourier transform respectively by

$$\|u\|_1 = \int_{\mathbb{R}} |u(x)| dx, \quad u^*(y) = \int_{\mathbb{R}} e^{iyx} u(x) dx, y \in \mathbb{R}. \quad (2.11)$$

When u, v are square integrable, we denote the \mathbb{L}^2 -norm and the \mathbb{L}^2 scalar product by

$$\|u\| = \left(\int_{\mathbb{R}} |u(x)|^2 dx \right)^{1/2}, \quad \langle u, v \rangle = \int_{\mathbb{R}} u(x) \bar{v}(x) dx \text{ with } z\bar{z} = |z|^2. \quad (2.12)$$

We recall that, for any integrable and square-integrable functions u, u_1, u_2 , the following relations hold:

$$(u^*)^*(x) = 2\pi u(-x) \text{ and } \langle u_1, u_2 \rangle = (2\pi)^{-1} \langle u_1^*, u_2^* \rangle. \quad (2.13)$$

The convolution product of u, v is denoted by:

$$u \star v(x) = \int_{\mathbb{R}} u(y) \bar{v}(x-y) dy.$$

3 Moment and small sample properties

For statistical purposes, the existence of moments of (L_t) is required. This is why we introduce the following assumption:

(H2-(l)) For l integer, $\int_{|x|>1} |x|^l n(x) dx < \infty$.

According to [58], Section 5.25, Theorem 5.23, $\mathbb{E}|L_t|^l < \infty$ is equivalent to (H2-(l)). Note that the integrability of $n(\cdot)$ near 0 is in all cases ruled by (2.3) and by Assumption (H1-g) in the finite variation case.

The following proposition relates the moments of $Z_1 = L_\Delta$ under (H2-(l)) to the integrals

$$m_l = \int_{\mathbb{R}} x^l n(x) dx = \int_{\mathbb{R}} x^l n(x) dx. \quad (3.1)$$

Proposition 3.1 1. Assume (H1-g) and (H2-(l)) with $l \geq 2$. Then, $\mathbb{E}(Z_1) = \Delta m_1$, $\mathbb{E}(Z_1^2) = \Delta m_2 + \Delta^2 m_1^2$, and more generally, for $2 \leq q \leq l$,

$$\mathbb{E}(Z_1^q) = \Delta m_q + o(\Delta).$$

2. Assume (H2-(l)) with $l \geq 2$. Then, $\mathbb{E}(Z_1) = \Delta b$, $\mathbb{E}(Z_1^2) = \Delta(\sigma^2 + m_2) + \Delta^2 b^2$. When $l \geq 3$ and $3 \leq q \leq l$,

$$\mathbb{E}(Z_1^q) = \Delta m_q + o(\Delta).$$

Proof. Assumption (H2-(l)) ensures the existence of moments up to order l in all cases.

Under (H1-g) and (H2-(l)), the characteristic exponent (2.5) is l times differentiable

with $\psi^{(j)}(0) = i^j m_j$ for $j \leq l$. Therefore, the j -th order cumulant of Z_1 is $\kappa_j = \Delta m_j$. Denoting by μ_j the j -th order moment of Z_1 , we have the classical relation between cumulants and moments:

$$\kappa_j = \mu_j - \sum_{i=1}^{j-1} \binom{j-1}{i-1} \kappa_i \mu_{j-i}. \quad (3.2)$$

We have $\kappa_1 = \mathbb{E}(Z_1)$, $\kappa_2 = \text{Var}(Z_1)$ and by elementary induction, we get the result for higher order moments.

In the general case, we derivate (2.9) to compute the cumulants of L_1 :

$$\psi'(0) = ib, \quad \psi''(0) = -(\sigma^2 + m_2), \quad \text{for } q \geq 3, \quad \psi^{(q)}(0) = i^q m_q.$$

The result follows.

The previous proposition shows that all moments of Z_1 are of order $O(\Delta)$.

We now look at absolute moments under different conditions.

Proposition 3.2 1. Assume (H2-(r)) and $r > 2$. Then,

$$\mathbb{E}|Z_1|^r = \Delta \int |x|^r n(x) dx + o(\Delta).$$

2. Assume (H1-g) and for $r \leq 1$, $\int |x|^r n(x) dx < \infty$. Then, $\mathbb{E}|Z_1|^r \leq \Delta \int |x|^r n(x) dx$.
3. Let $L_t = B_{\Gamma_t}$ where (Γ_t) is a pure jump increasing Lévy process (subordinator) with Lévy density n_Γ satisfying $\int_0^{+\infty} \gamma n_\Gamma(\gamma) d\gamma < \infty$ and (B_t) is a Brownian motion independent of (Γ_t) . The Lévy measure of (L_t) has a density given by

$$n(x) = \int_0^{+\infty} e^{-x^2/2\gamma} \frac{1}{\sqrt{2\pi\gamma}} n_\Gamma(\gamma) d\gamma. \quad (3.3)$$

If $c_r = \int_0^{+\infty} \gamma^{r/2} n_\Gamma(\gamma) d\gamma < \infty$ with $r \leq 2$, $\mathbb{E}|L_\Delta|^r \leq \Delta c_r C_r$, where $C_r = \mathbb{E}|X|^r$, for X a standard Gaussian variable.

4. Let (L_t) be a Lévy process with no Gaussian component. Then, $L_\Delta/\sqrt{\Delta}$ converges to 0 as Δ tends to 0 in probability and in \mathbb{L}^r for all $r < 2$.

Proof. For the first point, we refer to [30].

For the second point, the assumptions and the fact that $r \leq 1$ imply

$$|Z_1|^r = |L_\Delta|^r = \left| \sum_{s \leq \Delta} L_s - L_{s-} \right|^r \leq \sum_{s \leq \Delta} |L_s - L_{s-}|^r.$$

Taking expectations yields the result.

For the third point, consider f a non-negative function such that $f(0) = 0$. We have:

$$\mathbb{E} \sum_{s \leq t} f(L_s - L_{s-}) = \mathbb{E} \sum_{s \leq t} f(B_{\Gamma_s} - B_{\Gamma_{s-}}).$$

Thus, $\sum_{s \leq t} \mathbb{E} f(B_{\Gamma_s} - B_{\Gamma_{s-}}) = \sum_{s \leq t} \int_{\mathbb{R}} f(x) \mathbb{E} \left(e^{(-x^2/2(\Gamma_s - \Gamma_{s-}))} \frac{1}{\sqrt{2\pi(\Gamma_s - \Gamma_{s-})}} \right) dx$. For all x , we have

$$\mathbb{E} \sum_{s \leq t} \left(e^{(-x^2/2(\Gamma_s - \Gamma_{s-}))} \frac{1}{\sqrt{2\pi(\Gamma_s - \Gamma_{s-})}} \right) = t \int_0^{+\infty} e^{-x^2/2\gamma} \frac{1}{\sqrt{2\pi\gamma}} n_{\Gamma}(\gamma) d\gamma.$$

Therefore, we get the formula for the Lévy density n of (L_t) . Moreover,

$$\int_{\mathbb{R}} |x|^\alpha n(x) dx = C_\alpha \int_0^{+\infty} \gamma^{\alpha/2} n_{\Gamma}(\gamma) d\gamma.$$

Thus $\mathbb{E}|L_\Delta|^r = C_r \mathbb{E}(\Gamma_\Delta^{r/2})$. As $r/2 \leq 1$, $\Gamma_\Delta^{r/2} = (\sum_{s \leq \Delta} \Gamma_s - \Gamma_{s-})^{r/2} \leq \sum_{s \leq \Delta} (\Gamma_s - \Gamma_{s-})^{r/2}$. Taking expectation gives the result.

For the last point, we refer to [5] (Theorem 1, p. 804), see also [1].

Let us now look at small sample properties of the distribution of Z_1 .

Proposition 3.3 *Let P_Δ denote the distribution of Z_1 . Define*

$$\mu_\Delta^{(l)} = \Delta^{-1} x^l P_\Delta(dx), \quad \mu^{(l)}(dx) = x^l n(x) dx. \quad (3.4)$$

1. *Assume (H1-g). The distribution $\mu_\Delta^{(1)}$ has a density g_Δ given by*

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

and converges weakly to $\mu^{(1)}$ as Δ tends to 0.

2. *Under (H1-l), $\mu_\Delta^{(2)}$ converges weakly to $\mu^{(2)}$ as Δ tends to 0.*

3. *Under (H1-p), $\mu_\Delta^{(3)}$ converges weakly to $\mu^{(3)}$ as Δ tends to 0.*

Proof. Recall that $g(x) = x n(x)$. Under (H1-g),

$$\int \mathbb{E}|g(x - Z_1)| dx = \mathbb{E} \int |g(x - Z_1)| dx = \int |g(x)| dx < +\infty.$$

Thus $\mathbb{E}|g(x - Z_1)| < +\infty$ a.e. (dx) , which implies that $\mathbb{E}(g(x - Z_1))$ is a.e. well defined. Derivating φ_Δ and using (2.5) yields

$$\Delta^{-1} \varphi'_\Delta(u) = i \Delta^{-1} \mathbb{E}(Z_1 e^{iuZ_1}) = \varphi_\Delta(u) \psi'(u) \quad (3.5)$$

where

$$\psi'(u) = i g^*(u). \quad (3.6)$$

Therefore, the Fourier transforms of $\mu_\Delta^{(1)}$, $\mu^{(1)}$, P_Δ satisfy

$$(\mu_\Delta^{(1)})^* = (\mu^{(1)})^* P_\Delta^*.$$

Consequently, $\mu_\Delta^{(1)} = \mu^{(1)} \star P_\Delta$. This gives the result for the density of $\mu_\Delta^{(1)}$. The weak convergence is a consequence of the fact that $\varphi_\Delta(u)$ tends to 0 as Δ tends to 0. Under (H1- ℓ), derivating φ_Δ a second time yields

$$\Delta^{-1}\varphi_\Delta''(u) = i^2\Delta^{-1}\mathbb{E}(Z_1^2 e^{iuZ_1}) = \varphi_\Delta(u)\psi''(u) + \Delta\varphi_\Delta(u)(\psi'(u))^2 \quad (3.7)$$

Now using (2.7) and recalling that $\ell(x) = x^2n(x)$, we obtain:

$$\psi'(u) = i\left(b + \int_{\mathbb{R}}(e^{iux} - 1)x n(x)dx\right), \quad \psi''(u) = i^2\ell^*(u). \quad (3.8)$$

Therefore,

$$\Delta^{-1}\mathbb{E}(Z_1^2 e^{iuZ_1}) = -\Delta^{-1}\varphi_\Delta''(u) \rightarrow \ell^*(u).$$

Hence $\mu_\Delta^{(2)} \Longrightarrow \mu^{(2)}$ as $\Delta \rightarrow 0$.

Under (H1-p), derivating a third time φ_Δ , we get:

$$\begin{aligned} \Delta^{-1}\varphi_\Delta'''(u) &= i^3\Delta^{-1}\mathbb{E}(Z_1^3 e^{iuZ_1}) \\ &= \varphi_\Delta(u)\psi'''(u) + 3\Delta\varphi_\Delta(u)\psi'(u)\psi''(u) + \Delta^2\varphi_\Delta(u)(\psi'(u))^3 \end{aligned} \quad (3.9)$$

with, using (2.9) and $p(x) = x^3n(x)$,

$$\psi'(u) = i\left(b + iu\sigma^2 + \int_{\mathbb{R}}(e^{iux} - 1)x n(x)dx\right), \quad \psi''(u) = i^2(\sigma^2 + \ell^*(u)),$$

and

$$\psi'''(u) = i^3p^*(u). \quad (3.10)$$

This shows that

$$\Delta^{-1}\mathbb{E}(Z_1^3 e^{iuZ_1}) = i^{-3}\Delta^{-1}\varphi_\Delta'''(u) \rightarrow p^*(u).$$

Therefore, $\mu_\Delta^{(3)} \Longrightarrow \mu^{(3)}$ as $\Delta \rightarrow 0$.

Note that the Lévy measure can always be obtained as a limit: for every fixed $a > 0$, $(1/\Delta)P_\Delta(dx)$ converges vaguely on $|x| > a$ as $\Delta \rightarrow 0$ to $n(x)dx$, see e.g. [9], p. 39, ex. 5.1.

The following elementary proposition gives the rate of convergence to 0 of φ_Δ .

Proposition 3.4 1. Under (H1-g), we have:

$$|\varphi_\Delta(u) - 1| \leq |u|\Delta\|g\|_1. \quad (3.11)$$

2. If $\int_{\mathbb{R}}x^2n(x)dx < +\infty$,

$$|\varphi_\Delta(u) - 1| \leq \Delta|u|(c(u) + \sigma^2|u|)$$

where $c(u) = |b| + \left|\int_0^u|\ell^*(v)|dv\right|$. If ℓ^* is integrable on \mathbb{R} , then

$$|\varphi_\Delta(u) - 1| \leq \Delta |u| (|b| + \|\ell^*\|_1 + |u| \sigma^2). \quad (3.12)$$

Proof. By the Taylor formula,

$$\varphi_\Delta(u) - 1 = u \varphi'_\Delta(c_u u) = iu \Delta \varphi_\Delta(c_u u) \psi'(c_u u),$$

for some $c_u \in (0, 1)$.

Under (H1-g), $|\psi'(u)| = |g^*(u)| \leq \|g\|_1$ (see (3.6)). Inequality (3.11) follows.

For the second point, we use (3.10) and the relation $e^{iux} - 1 = ix \int_0^u e^{ivx} dv$ to obtain:

$$\psi'(u) = ib - u\sigma^2 - \int_{\mathbb{R}} \left(\int_0^u e^{ivx} dv \right) x^2 n(x) dx = ib - u\sigma^2 - \int_0^u \ell^*(v) dv.$$

This gives the two inequalities.

4 Adaptive estimation in the pure jump case

We consider now a Lévy process (L_t) discretely observed with sampling interval Δ under the asymptotic framework (2.1) and assume that (H1-g) holds and that the characteristic exponent is

$$\psi(u) = \int_{\mathbb{R}} (e^{iux} - 1) n(x) dx. \quad (4.1)$$

For the estimation of $g(x) = xn(x)$, (H1-g), (H2-(l)) for an integer l to be precised in each proposition or theorem and the following additional assumptions are required.

(H3-g) The function g belongs to $\mathbb{L}^2(\mathbb{R})$.

(H4-g) $M_2 := \int x^2 g^2(x) dx < +\infty$.

Assumptions (H1-g) and (H2-(l)) are moment assumptions for the *i.i.d.* observed random variables $(Z_k = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \dots, n)$ (see Section 3, Proposition 3.1). Under (H1-g), (H2-(l)) for $l > 1$ implies (H2-(k)) for $k \leq l$.

Noting that

$$\|g\|_1^2 := \left(\int |g(x)| dx \right)^2 \leq \int (1 + |x|)^2 g^2(x) dx \int \frac{dx}{(1 + |x|)^2},$$

we see that (H3-g)-(H4-g) imply (H1-g).

Let us describe the ideas on which rely the statistical strategies: estimation of g by a deconvolution approach, estimation of g on a compact subset of \mathbb{R} and kernel estimation of g .

4.1 Deconvolution approach

The first strategy is based on deconvolution. By (H1-g), derivating φ_Δ yields the following expression for the Fourier transform of g :

$$g^*(u) = -i\psi'(u) = -i\frac{\Delta^{-1}\varphi'_\Delta(u)}{\varphi_\Delta(u)}. \quad (4.2)$$

As the r.h.s. depends on the distribution of the observations, this relation suggests to estimate g^* and then build an estimator of g by Fourier inversion, thus relating the Lévy density estimation with deconvolution.

Let us make a short parenthesis to clarify the standard deconvolution problem. Suppose that observations $Y_i = X_i + \varepsilon_i, i = 1, \dots, n$ are available where the two samples (X_i) and (ε_i) are independent, composed of *i.i.d.* random variables, the X_i 's have density f_X and the ε_i 's have density f_ε . The random variables of interest are the X_i 's and the ε_i 's are an observation noise called observation error. If the Fourier transform of the noise distribution is never null, the relation

$$f_X^* = \frac{f_Y^*}{f_\varepsilon^*}$$

suggests to estimate the r.h.s. and deduce an estimator of f_X by Fourier inversion. A key distinction appears at this stage. Either the noise distribution is known (deconvolution with known errors distribution) or it is not (deconvolution with unknown errors distribution). The latter problem is clearly more difficult than the former. With known errors distribution, only the estimation of f_Y^* is required. This is usually done by using an empirical estimator. With unknown errors distribution, the estimation of f_ε^* is also required. This raises lots of difficulties. Detailed references are given and discussed in Section 11.

The link between deconvolution and estimation of g is now clear. Formula (4.2) shows that $g^*(u)$ is a quotient of two unknown Fourier transforms. The numerator is

$$\Delta^{-1}\theta_\Delta(u) := -i\Delta^{-1}\varphi'_\Delta(u) = \Delta^{-1}\mathbb{E}Z_k e^{iuZ_k} = g_\Delta^*(u), \quad (4.3)$$

where g_Δ is the density of the measure $\mu_\Delta^{(1)}$ (see Proposition (3.3)). The denominator $\varphi_\Delta(u)$ which is non null is the Fourier transform of the distribution P_Δ of Z_1 . Numerator and denominator being linked with the unknown distribution of Z_1 , we are faced with a problem closely related to deconvolution with unknown errors distributions. In the LF framework, numerator and denominator have to be estimated with the same sample (Z_k) . References are given in Section 11. The HF frequency framework provides a simplification. Indeed, as $\varphi_\Delta \rightarrow 1$, the estimation of the denominator becomes useless. The price to pay is an additional term which is a bias. Relation (4.2) may be written as:

$$-i\Delta^{-1}\varphi'_\Delta(u) = g^*(u) + g^*(u)(\varphi_\Delta(u) - 1) = \Delta^{-1}\mathbb{E}(Z_k e^{iuZ_k}) = \Delta^{-1}\theta_\Delta(u). \quad (4.4)$$

Simply using an empirical estimator of $\Delta^{-1}\varphi'_\Delta(u)$ yields an estimator of $g^*(u)$. Let us set

$$\hat{\theta}_\Delta(u) = \frac{1}{n} \sum_{k=1}^n Z_k e^{iuZ_k}, \quad \widehat{g^*(u)} = \Delta^{-1} \hat{\theta}_\Delta(u). \quad (4.5)$$

Note that, using Proposition 3.4, the bias of $\widehat{g^*(u)}$ as a pointwise estimator of $g^*(u)$ satisfies, under (H1-g),

$$|\mathbb{E}(\widehat{g^*(u)}) - g^*(u)| = |\Delta^{-1} \theta_\Delta(u) - g^*(u)| \leq |u| \Delta \|g\|_1^2. \quad (4.6)$$

The following inequalities are useful for the variance of the estimator $\widehat{g^*(u)}$.

Proposition 4.1 *Under (H1-g) and (H2-(2p)), for $p \geq 1$, there exists a constant C_p such that*

$$\mathbb{E}(|\widehat{g^*(u)} - \mathbb{E}(\widehat{g^*(u)})|^{2p}) \leq \frac{C_p}{(n\Delta)^p}. \quad (4.7)$$

Note that for $p = 1$, (4.7) is a simple variance inequality:

$$\mathbb{E}(|\widehat{g^*(u)} - \mathbb{E}(\widehat{g^*(u)})|^2) \leq \frac{1}{n\Delta} (m_2 + \Delta m_1^2) = \frac{1}{n\Delta^2} \mathbb{E}(Z_1^2). \quad (4.8)$$

Proof. For $p = 1$, (4.8) follows from:

$$\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2) = \frac{1}{n} \text{Var}(Z_1 \exp(iuZ_1)) \leq \frac{1}{n} \mathbb{E}(Z_1^2).$$

For $p \geq 1$, we apply Rosenthal's inequality recalled in Appendix (see (.1)):

$$\begin{aligned} \mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^{2p}) &\leq \frac{C(2p)}{n^{2p}} \left(\sum_{k=1}^n \mathbb{E}[|Z_k e^{iuZ_k} - \mathbb{E}(Z_k e^{iuZ_k})|^{2p}] \right. \\ &\quad \left. + \left(\sum_{k=1}^n \mathbb{E}[|Z_k e^{iuZ_k} - \mathbb{E}(Z_k e^{iuZ_k})|^2] \right)^p \right) \\ &\leq \frac{C'(2p)}{n^{2p}} (n \mathbb{E}(Z_1^{2p}) + n^p (\mathbb{E}(Z_1^2))^p). \end{aligned}$$

Dividing both sides by $(n\Delta)^{2p}$ and using that all moments have order Δ (Proposition 3.1), we get

$$\mathbb{E}(|\widehat{g^*(u)} - \mathbb{E}(\widehat{g^*(u)})|^{2p}) \leq C''(2p) \left(\frac{1}{(n\Delta)^{2p-1}} + \frac{1}{(n\Delta)^p} \right).$$

We conclude using that $p \geq 1$.

The following inequality for empirical moments holds.

Proposition 4.2 *Assume (H1-g). If pl is even, (H2-(pl)) and (H2-(2l)) hold, then there exists a constant C_p such that*

$$\mathbb{E} \left(\left| \frac{1}{n\Delta} \sum_{k=1}^n Z_k^l - \mathbb{E}(Z_1^l) \right|^p \right) \leq C_p \left(\frac{1}{(n\Delta)^{p-1}} + \frac{1}{(n\Delta)^{p/2}} \right). \quad (4.9)$$

The proof is almost identical to the proof of Proposition 4.1 with the use of Rosenthal's inequality and is omitted.

4.1.1 Definition of a collection of estimators

In this paragraph, we present a collection of estimators (\hat{g}_m) , indexed by a positive parameter m that will below be subject to constraints for adaptivity results. Distinct constructions give rise to this class of estimators, each having its own interest for interpretation, implementation or theoretical aspects. We start with the simple cut-off approach.

To build an estimator of g , we have at our disposal an estimator of g^* given by $\widehat{g^*} = \widehat{\theta}_\Delta / \Delta$ (see (4.5)). This function is not integrable so that we cannot simply take its inverse Fourier transform. The cut-off approach consists in introducing a parameter $m > 0$, the cut-off parameter, and setting:

$$\hat{g}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \widehat{g^*}(u) du. \quad (4.10)$$

This first step provides a collection of estimators $(\hat{g}_m)_{m>0}$. A second step treated below is to define a data-driven choice \hat{m} of m to build the final estimator $\hat{g}_{\hat{m}}$. A key feature of \hat{g}_m lies in the relation

$$\hat{g}_m^* = \widehat{g^*}(u) \mathbb{I}_{[-\pi m, \pi m]}(u). \quad (4.11)$$

A second interesting property of \hat{g}_m is that the integral (4.10) is explicit. Introducing

$$\phi(x) = \frac{\sin(\pi x)}{\pi x} \quad (\text{with } \phi(0) = 1), \quad (4.12)$$

a simple integration leads to

$$\hat{g}_m(x) = \frac{m}{n\Delta} \sum_{k=1}^n Z_k \phi(m(Z_k - x)).$$

Therefore \hat{g}_m may be interpreted as a kernel estimator with kernel ϕ and bandwidth $1/m$. Formula (4.10) allows to study the \mathbb{L}^2 -risk of \hat{g}_m for all m . We need to introduce

$$g_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} g^*(u) du,$$

which is such that

$$g_m^* = g^* \mathbb{I}_{[-\pi m, \pi m]} \quad \text{and} \quad (g - g_m)^* = g^* \mathbb{I}_{[-\pi m, \pi m]^c}. \quad (4.13)$$

Proposition 4.3 Assume that (H1-g)- (H2-(2))- (H3-g) hold. Then for all positive m ,

$$\mathbb{E}(\|g - \hat{g}_m\|^2) \leq \|g - g_m\|^2 + \mathbb{E}(Z_1^2/\Delta) \frac{m}{n\Delta} + \frac{\|g\|_1^2}{2\pi} \Delta^2 \int_{-\pi m}^{\pi m} u^2 |g^*(u)|^2 du.$$

Remark 4.1 In the above inequality, $\|g - g_m\|^2$ is a square bias which decreases with m , due to the estimation method: g_m is estimated instead of g . The second term bounds the variance of the estimator \hat{g}_m and increases with m . As a minimal condition to bound the variance term, we impose below $m \leq n\Delta$. The last term comes from the fact that we have neglected $g^*(u)(\varphi_\Delta(u) - 1)$ when building the estimator. It is a bias of the estimating method.

Proof. By the Parseval equality, $\|\hat{g}_m - g\|^2 = \|\hat{g}_m^* - g^*\|^2/(2\pi)$. Using definitions (4.5) and (4.3) yields

$$\begin{aligned} \mathbb{E}(\|\hat{g}_m - g\|^2) &= \frac{1}{2\pi} [\mathbb{E}(\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta}) \mathbb{I}_{[-\pi m, \pi m]}\|^2) + \|(\frac{\theta_\Delta}{\Delta} - g^*) \mathbb{I}_{[-\pi m, \pi m]} - g^* \mathbb{I}_{[-\pi m, \pi, m]^c}\|^2] \\ &\leq \frac{1}{2\pi} \left[\mathbb{E}(\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta}) \mathbb{I}_{[-\pi m, \pi m]}\|^2) + \|(\frac{\theta_\Delta}{\Delta} - g^*) \mathbb{I}_{[-\pi m, \pi m]}\|^2 \right] \\ &\quad + \frac{1}{2\pi} \|g^* \mathbb{I}_{[-\pi m, \pi, m]^c}\|^2. \end{aligned}$$

By (4.13) and the Parseval equality, the last term is exactly $\|g - g_m\|^2$. For the second term, using (4.4) and (3.11), we have

$$\|(\frac{\theta_\Delta}{\Delta} - g^*) \mathbb{I}_{[-\pi m, \pi m]}\|^2 = \|(\varphi_\Delta - 1)g^* \mathbb{I}_{[-\pi m, \pi m]}\|^2 \leq \Delta^2 \|g\|_1^2 \int_{-\pi m}^{\pi m} u^2 |g^*(u)|^2 du.$$

Lastly, (4.8) yields

$$\mathbb{E}(\|(\frac{\hat{\theta}_\Delta}{\Delta} - \frac{\theta_\Delta}{\Delta}) \mathbb{I}_{[-\pi m, \pi m]}\|^2) = \int_{-\pi m}^{\pi m} \Delta^{-2} \mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2) du \leq \frac{2\pi m \mathbb{E}(Z_1^2)}{n\Delta^2}.$$

By gathering the three bounds, we obtain the result.

4.1.2 Rates of convergence

Rates of convergence of the \mathbb{L}^2 -risk can be deduced from Proposition 4.3. In deconvolution, the regularity classes for rates interpretation are usually Sobolev classes such as

$$\mathcal{C}(a, L) = \left\{ g \in (\mathbb{L}^1 \cap \mathbb{L}^2)(\mathbb{R}), \int (1+u^2)^a |g^*(u)|^2 du \leq L \right\}. \quad (4.14)$$

The following holds:

Proposition 4.4 Assume that (H1-g)- (H2-(2))- (H3-g) hold and that g belongs to $\mathcal{C}(a, L)$. Assume that $m \leq n\Delta$ and in addition to the asymptotic framework (2.1), that $n\Delta^2 \leq 1$. The following rate is obtained by choosing $m = O((n\Delta)^{1/(2a+1)})$:

$$\mathbb{E}(\|g - \hat{g}_m\|^2) \leq O((n\Delta)^{-2a/(2a+1)}).$$

If $a \geq 1$, then it is enough to have $n\Delta^3 = O(1)$ (instead of $n\Delta^2 \leq 1$).

Proof. We evaluate the infimum over m of the risk bound of Proposition 4.3. By relation (4.13), as $g \in \mathcal{C}(a, L)$, we get

$$\|g - g_m\|^2 = \frac{1}{2\pi} \int_{|u| \geq \pi m} |g^*(u)|^2 du \leq \frac{L}{2\pi} (\pi m)^{-2a}.$$

The optimal compromise between $\|g - g_m\|^2$ and $m/(n\Delta)$, infimum over m of the sum

$$\|g - g_m\|^2 + m/(n\Delta),$$

i.e. the first two terms in the risk bound of Proposition 4.3), is obtained for $m^{-2a} \propto m/(n\Delta)$, i.e. $m = O((n\Delta)^{1/(2a+1)})$ and leads to the rate $(n\Delta)^{-2a/(2a+1)}$.

We now look for a condition on Δ implying that the term $\Delta^2 \int_{-\pi m}^{\pi m} u^2 |g^*(u)|^2 du$ has order less than $(n\Delta)^{-2a/(2a+1)}$.

As $g \in \mathcal{C}(a, L)$,

$$\int_{-\pi m}^{\pi m} u^2 |g^*(u)|^2 du \leq Lm^{2(1-a)}.$$

If $a \geq 1$, the condition $\Delta^2 = O(1/(n\Delta))$, i.e. $n\Delta^3 = O(1)$ implies:

$$\Delta^2 \int_{-\pi m}^{\pi m} u^2 |g^*(u)|^2 du = O(1/(n\Delta))$$

which is negligible. The risk bound order is $O((n\Delta)^{-2a/(2a+1)})$.

If $a \in (0, 1)$, we must have at least $\Delta^2 m^{2(1-a)} \leq m^{-2a}$. Hence, $\Delta^2 m^2 \leq 1$. This is achieved for $n\Delta^2 \leq 1$ as $m \leq n\Delta$. The risk bound order is again $O((n\Delta)^{-2a/(2a+1)})$.

Remark 4.2 If $n\Delta^2 \leq 1$ and if g is analytic i.e. belongs to a class

$$\mathcal{A}(\gamma, \mathcal{Q}) = \{f, \int (e^{\gamma x} + e^{-\gamma x})^2 |f^*(x)|^2 dx \leq \mathcal{Q}\},$$

then the risk is of order $O(\log(n\Delta)/(n\Delta))$ (choose $m = O(\log(n\Delta))$).

4.1.3 Adaptive estimator

In this paragraph, the selection method of a relevant data-driven cut-off parameter m is described. The choice should lead to an adaptive estimator. An estimator is adaptive if its \mathbb{L}^2 -risk attains automatically the best possible rate of convergence to 0 without any knowledge of the regularity of g .

For this, it is convenient to use the property that the estimators \hat{g}_m are projection estimators, obtained as minimizers of a projection contrast. For positive m , consider the following closed subspace of $\mathbb{L}^2(\mathbb{R})$

$$S_m = \{t \in \mathbb{L}^2(\mathbb{R}), \text{supp}(t^*) \subset [-\pi m, \pi m]\}. \quad (4.15)$$

Let us give the main properties of the collection of spaces (S_m) . For $t \in \mathbb{L}^2(\mathbb{R})$, let t_m denote its orthogonal projection on S_m . The function t_m is characterized by the fact that

$$t_m^* = t^* \mathbb{I}_{[-\pi m, \pi m]}.$$

Hence,

$$\|t - t_m\|^2 = \frac{1}{2\pi} \|t^* - t_m^*\|^2 = \frac{1}{2\pi} \int_{|x| \geq \pi m} |t^*(x)|^2 dx.$$

The function g_m defined above is thus the orthogonal projection of g on S_m and \hat{g}_m belongs to S_m (see (4.11) and (4.13)).

Moreover, for $t \in S_m$, $t(x) = (1/2\pi) \int_{-\pi m}^{\pi m} e^{-iux} t^*(u) du$, and

$$|t(x)| \leq \frac{1}{2\pi} \left(\int_{-\pi m}^{\pi m} |t^*(u)|^2 du \int_{-\pi m}^{\pi m} |e^{iux}|^2 du \right)^{1/2}.$$

Thus

$$\forall t \in S_m, \|t\|_\infty := \sup_{x \in \mathbb{R}} |t(x)| \leq \sqrt{m} \|t\|. \quad (4.16)$$

Let, for $t \in S_m$,

$$\begin{aligned} \gamma_n(t) &= \|t\|^2 - \frac{1}{\pi} \int \frac{\hat{\theta}_\Delta(u)}{\Delta} t^*(-u) du = \|t\|^2 - \frac{2}{2\pi} \langle \hat{g}_m^*, t^* \rangle \\ &= \|t\|^2 - 2 \langle \hat{g}_m, t \rangle = \|t - \hat{g}_m\|^2 - \|\hat{g}_m\|^2. \end{aligned} \quad (4.17)$$

Evidently,

$$\hat{g}_m = \arg \min_{t \in S_m} \gamma_n(t),$$

and

$$\gamma_n(\hat{g}_m) = -\|\hat{g}_m\|^2.$$

Using (4.10) and (4.12), we have

$$\|\hat{g}_m\|^2 = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{\theta}_\Delta(u)}{\Delta} \right|^2 du = \frac{m}{n^2 \Delta^2} \sum_{1 \leq k, l \leq n} Z_k Z_l \phi(m(Z_k - Z_l)). \quad (4.18)$$

Finally, it is interesting to stress that the space S_m is generated by an orthonormal basis, the sinus cardinal basis, given by:

$$\phi_{m,j}(x) = \sqrt{m} \phi(mx - j), \quad j \in \mathbb{Z} \quad (4.19)$$

where ϕ is defined by (4.12) (see [54], p.22). This can be seen noting that:

$$\phi_{m,j}^*(x) = \frac{e^{ixj/m}}{\sqrt{m}} \mathbb{I}_{[-\pi m, \pi m]}(x). \quad (4.20)$$

As above, we use that $\phi_{m,j}(x) = (1/2\pi) \int_{-\pi m}^{\pi m} e^{iux} \phi_{m,j}^*(-u) du$ to obtain

$$\sum_{j \in \mathbb{Z}} \phi_{m,j}^2(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |e^{iux}|^2 du = m.$$

For $f \in \mathbb{L}^2(\mathbb{R})$, its orthogonal projection f_m on S_m can be written as

$$f_m = \sum_{j \in \mathbb{Z}} a_{m,j}(f) \phi_{m,j} \quad \text{with } a_{m,j}(f) = \langle f, \phi_{m,j} \rangle.$$

This leads to a third formulation of \hat{g}_m :

$$\hat{g}_m = \sum_{j \in \mathbb{Z}} \hat{a}_{m,j} \phi_{m,j} \quad \text{where } \hat{a}_{m,j} = \frac{1}{2\pi\Delta} \int \hat{\theta}_\Delta(u) \phi_{m,j}^*(-u) du = \frac{1}{n\Delta} \sum_{k=1}^n Z_k \phi_{m,j}(Z_k).$$

Using the development of \hat{g}_m on the orthonormal basis $(\phi_{m,j})_j$, we have

$$\|\hat{g}_m\|^2 = \sum_{j \in \mathbb{Z}} |\hat{a}_{m,j}|^2.$$

Although S_m is infinite-dimensional, we need not truncate the series to compute \hat{g}_m and $\|\hat{g}_m\|^2$ as we can use the explicit formulae (4.10) and (4.18). This is important for practical implementation. Nevertheless, the introduction of the basis is crucial for the proof.

We consider a collection $(S_m, m = 1, \dots, m_n)$ where m_n is restricted to satisfy $m_n \leq n\Delta$ and set

$$\hat{m} = \arg \min_{m \in \{1, \dots, m_n\}} (\gamma_n(\hat{g}_m) + \text{pen}(m)) \quad \text{with } \text{pen}(m) = \kappa \left(\frac{1}{n\Delta} \sum_{k=1}^n Z_k^2 \right) \frac{m}{n\Delta}.$$

We shall denote by

$$\text{pen}_{th}(m) = \mathbb{E}(\text{pen}(m)) = \kappa(\mathbb{E}(Z_1^2)/\Delta) \frac{m}{n\Delta}.$$

The intuition behind the selection criterion is the following. The risk can be decomposed in two terms:

$$\|g - \hat{g}_m\|^2 = \|g - g_m\|^2 + \|g_m - \hat{g}_m\|^2.$$

The \mathbb{L}^2 -orthogonality of the two terms is due to the disjoint supports of their Fourier transforms. To define the data-driven criterion, we replace the terms of the sum by

estimators. For the first term which is the bias, we have $\|g - g_m\|^2 = \|g\|^2 - \|g_m\|^2$. Noting that $\gamma_n(\hat{g}_m) = -\|\hat{g}_m\|^2$, $\gamma_n(\hat{g}_m)$ is up to a constant an estimator of the bias. The variance term $\mathbb{E}(\|g_m - \hat{g}_m\|^2)$ is estimated by $\text{pen}(m)$ where the constant κ is a numerical value to be tuned to avoid under-penalization (see Proposition 4.3). The value \hat{m} realizes the best compromise between estimated bias and estimated variance terms.

The following theorem shows the adaptivity property of the estimator $\hat{g}_{\hat{m}}$.

Theorem 4.1 *Assume that (H2-(8))-(H3-g)-(H4-g) are fulfilled, that the asymptotic framework (2.1) holds and that $m_n \leq n\Delta$. Then there exists a universal constant κ such that*

$$\begin{aligned} \mathbb{E}(\|g - \hat{g}_{\hat{m}}\|^2) &\leq C \inf_{m \in \{1, \dots, m_n\}} (\|g - g_m\|^2 + \text{pen}_{rh}(m)) \\ &\quad + \frac{C'\Delta^2}{2\pi} \int_{-\pi m_n}^{\pi m_n} u^2 |g^*(u)|^2 du + \frac{C'' \log^2(n\Delta)}{n\Delta}. \end{aligned}$$

The calibration of the constant κ is a classical difficulty in such penalized methods. Most often, κ calibrated by numerical simulations (see Section 8).

In what sense is $\hat{g}_{\hat{m}}$ adaptive? The property is contained in the infimum term of the risk bound. Suppose that g belongs to a Sobolev regularity class $\mathcal{C}(a, L)$, with unknown a and L . In Proposition 4.4, it is proved that:

$$\inf_{m \in \{1, \dots, m_n\}} \left(\|g - g_m\|^2 + \frac{m}{n\Delta} \right) \leq C(n\Delta)^{-2a/(2a+1)}$$

and that

$$\int_{-\pi m_n}^{\pi m_n} u^2 |g^*(u)|^2 du \leq C\Delta^2 m_n^{2(1-a)_+},$$

for some constant C . Thus, the estimator is automatically (for some other constant C) such that

$$\mathbb{E}(\|g - \hat{g}_{\hat{m}}\|^2) \leq C \left[(n\Delta)^{-2a/(2a+1)} + \Delta^2 m_n^{2(1-a)_+} \right] + \frac{C'' \log^2(n\Delta)}{n\Delta}.$$

If either $(a \geq 1, n\Delta^3 = O(1))$ or $(0 < a < 1 \text{ and } n\Delta^2 = O(1))$, then

$$\mathbb{E}(\|g - \hat{g}_{\hat{m}}\|^2) = O((n\Delta)^{-2a/(2a+1)}).$$

This rate is obtained without requiring the knowledge of a nor L in the procedure.

4.1.4 Proof of Theorem 4.1

To deal with the randomness of the penalty $\text{pen}(m)$, the proof is given in two steps. We define, for some $b, 0 < b < 1$,

$$\Omega_b := \left\{ \left| \frac{(1/n\Delta) \sum_{k=1}^n Z_k^2}{\mathbb{E}(Z_1^2/\Delta)} - 1 \right| \leq b \right\}, \quad (4.21)$$

so that $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2) = \mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b}) + \mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b^c})$.

Step 1. Study of $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b})$. By (4.17), we can write

$$\gamma_n(t) - \gamma_n(s) = \|t - g\|^2 - \|s - g\|^2 - 2\langle t - s, \hat{g}_m \rangle. \quad (4.22)$$

For $t \in S_m$, let us introduce the linear processes:

$$v_n(t) = \frac{1}{2\pi} \int \frac{\hat{\theta}_\Delta(u) - \theta_\Delta(u)}{\Delta} t^*(-u) du = \langle \hat{g}_m - \mathbb{E}\hat{g}_m, t \rangle, \quad (4.23)$$

$$R_n(t) = \frac{1}{2\pi} \int (\varphi_\Delta(u) - 1) g^*(u) t^*(-u) du = \langle \mathbb{E}\hat{g}_m - g, t \rangle. \quad (4.24)$$

The contrast $\gamma_n(t)$ admits the following decomposition :

$$\gamma_n(t) - \gamma_n(s) = \|t - g\|^2 - \|s - g\|^2 - 2v_n(t - s) - 2R_n(t - s), \quad (4.25)$$

Note that $v_n = \bar{v}_n$ and $R_n = \bar{R}_n$ so that they are both real valued.

With a constant k_n to be given later on, define

$$\theta_\Delta^{(1)}(u) = \mathbb{E} \left(Z_1 \mathbb{I}_{(|Z_1| \leq k_n \sqrt{\Delta})} e^{iuZ_1} \right), \quad \theta_\Delta^{(2)}(u) = \mathbb{E} \left(Z_1 \mathbb{I}_{(|Z_1| > k_n \sqrt{\Delta})} e^{iuZ_1} \right) \quad (4.26)$$

and their empirical counterparts

$$\hat{\theta}_\Delta^{(1)}(u) = \frac{1}{n} \sum_{k=1}^n Z_k \mathbb{I}_{(|Z_k| \leq k_n \sqrt{\Delta})} e^{iuZ_k}, \quad \hat{\theta}_\Delta^{(2)}(u) = \frac{1}{n} \sum_{k=1}^n Z_k \mathbb{I}_{(|Z_k| > k_n \sqrt{\Delta})} e^{iuZ_k}. \quad (4.27)$$

We split v_n into $v_n^{(1)} + v_n^{(2)}$ with

$$v_n^{(1)}(t) = \frac{1}{2\pi\Delta} \int (\hat{\theta}_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)) t^*(-u) du,$$

and

$$v_n^{(2)}(t) = \frac{1}{2\pi\Delta} \int (\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)) t^*(-u) du.$$

The definition of $\hat{g}_{\hat{m}}$ implies that

$$\gamma_n(\hat{g}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(g_m) + \text{pen}(m) \quad (4.28)$$

where we recall that g_m denotes the orthogonal projection of g on S_m .

Using (4.25)-(4.28) yields that, for all $m = 1, \dots, m_n$,

$$\begin{aligned} \|\hat{g}_{\hat{m}} - g\|^2 &\leq \|g - g_m\|^2 + \text{pen}(m) + 2v_n^{(1)}(g_m - \hat{g}_{\hat{m}}) - \text{pen}(\hat{m}) \\ &\quad + 2R_n(g_m - \hat{g}_{\hat{m}}) + 2v_n^{(2)}(g_m - \hat{g}_{\hat{m}}) \end{aligned}$$

For $\rho_n = v_n^{(1)}, v_n^{(2)}, R_n$, we can write

$$2\rho_n(g_m - \hat{g}_{\hat{m}}) \leq 2\|g_m - \hat{g}_{\hat{m}}\| \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} |\rho_n(t)|.$$

Then, we use

$$2xy \leq \frac{1}{8}x^2 + 8y^2$$

and the fact that $S_m + S_{\hat{m}} \subset S_{m_m}$ to obtain

$$\begin{aligned} \|\hat{g}_{\hat{m}} - g\|^2 &\leq \|g - g_m\|^2 + \text{pen}(m) + \frac{3}{8}\|g_m - \hat{g}_{\hat{m}}\|^2 + 8 \sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - \text{pen}(\hat{m}) \\ &\quad + 8 \sup_{t \in S_{m_m}, \|t\|=1} [R_n(t)]^2 + 8 \sup_{t \in S_{m_m}, \|t\|=1} [v_n^{(2)}(t)]^2, \end{aligned}$$

$$\begin{aligned} \|\hat{g}_{\hat{m}} - g\|^2 &\leq \left(1 + \frac{3}{4}\right)\|g - g_m\|^2 + \text{pen}(m) + \frac{3}{4}\|\hat{g}_{\hat{m}} - g\|^2 \\ &\quad + 8 \left(\sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, \hat{m}) \right) + 8p(m, \hat{m}) - \text{pen}(\hat{m}) \\ &\quad + 8 \sup_{t \in S_{m_m}, \|t\|=1} [R_n(t)]^2 + 8 \sup_{t \in S_{m_m}, \|t\|=1} [v_n^{(2)}(t)]^2. \end{aligned}$$

The function $p(m, m')$ plugged in the last inequality is fixed in the following Lemma.

Lemma 4.1 *Under the Assumptions of Theorem 4.1, define*

$$p(m, m') = 4\mathbb{E}(Z_1^2 / \Delta) \frac{m \vee m'}{n\Delta}, \quad (4.29)$$

then, there exists a constant k such that for $k_n = k\sqrt{n}/\log n\Delta$,

$$\mathbb{E} \left(\sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, \hat{m}) \right)_+ \leq \frac{C}{n\Delta},$$

where C is a constant.

Before giving the proof of this Lemma, we finish Step 1. On Ω_b , the following inequality holds (by bounding the indicator by 1), for any choice of κ :

$$\forall m, (1 - b)\text{pen}_{th}(m) \leq \text{pen}(m) \leq (1 + b)\text{pen}_{th}(m). \quad (4.30)$$

Therefore

$$\begin{aligned}
\frac{1}{4}\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b} &\leq \frac{7}{4}\|g - g_m\|^2 + (1+b)\text{pen}_{th}(m) \mathbb{I}_{\Omega_b} \\
&+ 8 \left(\sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, \hat{m}) \right)_+ \\
&+ (8p(m, \hat{m}) - (1-b)\text{pen}_{th}(\hat{m})) \mathbb{I}_{\Omega_b} \\
&+ 8 \sup_{t \in S_{m_n}, \|t\|=1} [R_n(t)]^2 + 8 \sup_{t \in S_{m_n}, \|t\|=1} [v_n^{(2)}(t)]^2.
\end{aligned}$$

The constant κ is now chosen such that

$$\forall m, m' \in \{1, \dots, m_n\}, \quad 8p(m, m') \leq (1-b)(\text{pen}_{th}(m) + \text{pen}_{th}(m')),$$

that is $\kappa \geq 32/(1-b)$. In view of (4.29), this gives the choices

$$\text{pen}_{th}(m) = \frac{32}{1-b} \mathbb{E}(Z_1^2/\Delta) \frac{m}{n\Delta} \quad \text{and} \quad \text{pen}(m) = \frac{32}{1-b} \frac{1}{n\Delta} \sum_{i=1}^n Z_i^2 \frac{m}{n\Delta}.$$

It follows that

$$\begin{aligned}
\frac{1}{4}\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b} &\leq \frac{7}{4}\|g - g_m\|^2 + 2\text{pen}_{th}(m) \\
&+ 8 \left(\sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, \hat{m}) \right)_+ \\
&+ 8 \sup_{t \in S_{m_n}, \|t\|=1} [R_n(t)]^2 + 8 \sup_{t \in S_{m_n}, \|t\|=1} [v_n^{(2)}(t)]^2.
\end{aligned}$$

Using (4.24) and (3.11), we get

$$\sup_{t \in S_{m_n}, \|t\|=1} R_n^2(t) \leq C\Delta^2 \int_{-\pi m_n}^{\pi m_n} u^2 |g^*(u)|^2 du. \quad (4.31)$$

For $v_n^{(2)}(t)$, we write

$$\begin{aligned}
\mathbb{E} \left(\sup_{t \in S_{m_n}, \|t\|=1} [v_n^{(2)}(t)]^2 \right) &\leq \frac{1}{2\pi\Delta^2} \int_{-\pi m_n}^{\pi m_n} \mathbb{E} |\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)|^2 du \\
&\leq \frac{\mathbb{E}(Z_1^2 \mathbb{I}_{|Z_1| > k_n \sqrt{\Delta}}) m_n}{n\Delta^2} \\
&\leq \frac{\mathbb{E}(Z_1^4) m_n}{nk_n^2 \Delta^3} = \frac{[\mathbb{E}(Z_1^4)/\Delta] m_n}{nk_n^2 \Delta^2} \leq \frac{[\mathbb{E}(Z_1^4)/\Delta]}{k_n^2 \Delta}
\end{aligned}$$

since $m_n \leq n\Delta$. We know that $[\mathbb{E}(Z_1^4)/\Delta]$ is bounded. If $k_n^2 \geq Cn/\log^2(n\Delta)$, then the above term is of order $\log^2(n\Delta)/(n\Delta)$. With the choice of $k_n = k\sqrt{n}/\log n\Delta$ for

some constant k prescribed by Lemma 4.1, the proof is achieved.

Step 1 can be concluded now. For all $m \in \{1, \dots, m_n\}$,

$$\begin{aligned} \mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b}) &\leq 7\|g - g_m\|^2 + 8\text{pen}_{th}(m) + \frac{C_1}{n\Delta} \\ &\quad + C_2\Delta^2 \int_{-\pi m_n}^{\pi m_n} u^2 |g^*(u)|^2 du + C_3 \frac{\log^2(n\Delta)}{n\Delta}. \end{aligned}$$

Proof of Lemma 4.1. We start by noting that

$$\mathbb{E}\left(\sup_{t \in S_m + S_{\hat{m}}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, \hat{m})\right)_+ \leq \sum_{m'=1}^{m_n} \mathbb{E}\left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, m')\right)_+.$$

For $t \in S_{m \vee m'} = S_m + S_{m'}$, $v_n^{(1)}(t)$ can be written as

$$v_n^{(1)}(t) = \frac{1}{n} \sum_{k=1}^n (f_t(Z_k) - \mathbb{E}(f_t(Z_k))),$$

where

$$f_t(z) = \frac{z \mathbb{I}_{|z| \leq k_n \sqrt{\Delta}}}{2\pi\Delta} \int_{-\pi(m \vee m')}^{\pi(m \vee m')} e^{ixz} t^*(-x) dx.$$

We intend to apply the Talagrand inequality (see Appendix) to the class

$$\mathcal{F} = \{f_t, t \in S_m + S_{m'}, \|t\| = 1\}.$$

We have to find the three quantities M , H , v .

Let $m'' = m \vee m'$. For $t \in S_{m''}$, using Inequality (4.16), we obtain

$$\sup_{z \in \mathbb{R}} |f_t(z)| \leq \frac{k_n}{2\pi\sqrt{\Delta}} \sup_{z \in \mathbb{R}} |2\pi t(z)| \leq \frac{k_n \|t\|_\infty}{\sqrt{\Delta}} \leq \frac{k_n \sqrt{m''}}{\sqrt{\Delta}} := M.$$

Clearly,

$$\mathbb{E}\left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2\right) \leq \frac{1}{2\pi\Delta^2} \int_{-\pi m''}^{\pi m''} \mathbb{E}|\hat{\theta}_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)|^2 du \leq \frac{\mathbb{E}(Z_1^2) m''}{n\Delta^2}.$$

Thus we set

$$H^2 = \frac{\mathbb{E}(Z_1^2) m''}{n\Delta^2}.$$

The most delicate term is v .

$$\begin{aligned} \text{Var}(f_t(Z_1)) &\leq \frac{1}{4\pi^2\Delta^2} \mathbb{E}\left(\iint Z_1^2 \mathbb{I}_{|Z_1| \leq k_n \sqrt{\Delta}} e^{i(x-y)Z_1} t^*(-x) t^*(y) dx dy\right) \\ &= \frac{1}{4\pi^2\Delta^2} \iint p_\Delta^*(x-y) t^*(-x) t^*(y) dx dy, \end{aligned}$$

where

$$p_{\Delta}^*(x) = \mathbb{E}(Z_1^2 \mathbb{I}_{|Z_1| \leq k_n \sqrt{\Delta}} e^{ixZ_1}).$$

Using that $t = \sum_{j \in \mathbb{Z}} t_j \phi_{m'',j}$ with $\|t\|^2 = \sum_{j \in \mathbb{Z}} t_j^2 = 1$,

$$\begin{aligned} \text{Var}(f_t(Z_1)) &\leq \frac{1}{4\pi^2 \Delta^2} \sum_{j,k \in \mathbb{Z}} t_j t_k \iint p_{\Delta}^*(x-y) \phi_{m'',j}^*(-x) \phi_{m'',k}^*(y) dx dy \\ &\leq \frac{1}{4\pi^2 \Delta^2} \left(\sum_{j,k \in \mathbb{Z}} \left| \iint p_{\Delta}^*(x-y) \phi_{m'',j}^*(-x) \phi_{m'',k}^*(y) dx dy \right|^2 \right)^{1/2}, \end{aligned}$$

Now, using Proposition 3.3, we have

$$p_{\Delta}^*(x) = \Delta \int z \mathbb{I}_{|z| \leq k_n \sqrt{\Delta}} e^{ixz} \mathbb{E}(g(z - Z_1)) dz.$$

This implies that (see (H4-g))

$$\begin{aligned} \int |p_{\Delta}^*(z)|^2 dz &\leq 2\pi \int |p_{\Delta}(z)|^2 dz = 2\pi \Delta^2 \int z^2 \mathbb{I}_{|z| \leq k_n \sqrt{\Delta}} \mathbb{E}^2(g(z - Z_1)) dz \\ &\leq 2\pi \Delta^2 \mathbb{E} \left(\int z^2 \mathbb{I}_{|z| \leq k_n \sqrt{\Delta}} g^2(z - Z_1) dz \right) \\ &\leq 4\pi \Delta^2 \mathbb{E} \left(\int (x^2 + Z_1^2) g^2(x) dx \right) = 4\pi \Delta^2 (M_2 + \mathbb{E}(Z_1^2) \|g\|^2). \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}(f_t(Z_1)) &\leq \frac{1}{4\pi^2 \Delta^2} \left(\iint_{[-\pi m'', \pi m'']^2} |p_{\Delta}^*(x-y)|^2 dx dy \right)^{1/2} \\ &\leq \frac{1}{4\pi^2 \Delta^2} (2\pi m'')^{1/2} \left(\int |p_{\Delta}^*(z)|^2 dz \right)^{1/2} \\ &\leq \frac{\sqrt{m''}}{\sqrt{2\pi} \Delta} (M_2 + \|g\|^2 \mathbb{E}(Z_1^2))^{1/2} := \nu. \end{aligned}$$

Applying Lemma .1 yields, for $\varepsilon^2 = 1/2$ and $p(m, m')$ given by (4.29),

$$\mathbb{E} \left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, m') \right)_+ \leq C_1 \left(\frac{\sqrt{m''}}{n\Delta} e^{-C_2 \sqrt{m''}} + \frac{k_n^2 m''}{n^2 \Delta} e^{-C_3 \sqrt{n}/k_n} \right)$$

as $p(m, m') = 4H^2$. We choose

$$k_n = k \frac{\sqrt{n}}{\log(n\Delta)} \quad \text{with} \quad k = \frac{C_3}{4}$$

and as $m \leq n\Delta$, we get

$$\mathbb{E} \left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, m') \right)_+ \leq C'_1 \left(\frac{\sqrt{m''}}{n\Delta} e^{-C_2 \sqrt{m''}} + \frac{1}{(\Delta n)^4 \log^2(n\Delta)} \right).$$

Therefore

$$\sum_{m'=1}^{m_n} \mathbb{E} \left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, m') \right)_+ \leq C'_1 \left(\frac{\sum_{m'=1}^{m_n} \sqrt{m''} e^{-C_2 \sqrt{m''}}}{n\Delta} + \frac{1}{(n\Delta)^3 \log^2(n\Delta)} \right).$$

As $C_2 x e^{-C_2 x}$ is decreasing for $x \geq 1/C_2$, and its maximum is $1/(eC_2)$, we get

$$\begin{aligned} \sum_{m'=1}^{m_n} \sqrt{m''} e^{-C_2 \sqrt{m''}} &\leq \sum_{\sqrt{m''} \leq 1/C_2} (eC_2)^{-1} + \sum_{\sqrt{m''} \geq 1/C_2} \sqrt{m''} e^{-C_2 \sqrt{m''}} \\ &\leq \frac{1}{eC_2^3} + \sum_{m'=1}^{\infty} \sqrt{m'} e^{-C_2 \sqrt{m'}} < +\infty. \end{aligned}$$

It follows that

$$\sum_{m'=1}^{m_n} \mathbb{E} \left(\sup_{t \in S_m + S_{m'}, \|t\|=1} [v_n^{(1)}(t)]^2 - p(m, m') \right)_+ \leq \frac{C}{n\Delta}$$

and Proposition 4.1 is proved. \square

Step 2. Study of $\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b^c})$.

This part is simpler. Using (4.25) and (4.28) yields that, $\forall m \in \{1, \dots, m_n\}$,

$$\begin{aligned} \|\hat{g}_{\hat{m}} - g\|^2 &\leq \|g - g_m\|^2 + \text{pen}(m) + 2v_n(g_m - \hat{g}_{\hat{m}}) - \text{pen}(\hat{m}) + 2R_n(g_m - \hat{g}_{\hat{m}}) \\ &\leq \|g - g_m\|^2 + \text{pen}(m) + \frac{1}{4}\|g_m - \hat{g}_{\hat{m}}\|^2 \end{aligned} \quad (4.32)$$

$$+ 8 \sup_{t \in S_{m_n}, \|t\|=1} [v_n(t)]^2 + 8 \sup_{t \in S_{m_n}, \|t\|=1} [R_n(t)]^2. \quad (4.33)$$

Now we apply inequality (4.31) to $R_n(t)$ and the Parseval formula for $v_n(t)$, and get

$$\begin{aligned} \frac{1}{2}\|\hat{g}_{\hat{m}} - g\|^2 &\leq \frac{3}{2}\|g - g_m\|^2 + \mathbb{E}(\text{pen}(m)) + [\text{pen}(m) - \mathbb{E}(\text{pen}(m))] \\ &\quad + \frac{4}{\pi\Delta^2} \int_{-\pi m_n}^{\pi m_n} |\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2 du + C'\Delta^2 \int_{-\pi m_n}^{\pi m_n} u^2 |g^*(u)|^2 du. \end{aligned}$$

Using that $\text{pen}_{th}(m) = \mathbb{E}(\text{pen}(m))$, we apply the Cauchy-Schwarz inequality and get:

$$\mathbb{E} \left((\text{pen}(m) - \text{pen}_{th}(m)) \mathbb{I}_{\Omega_b^c} \right) \leq \left\{ \mathbb{E} \left[\left(\frac{1}{n\Delta} \sum_{k=1}^n (Z_k^2 - \mathbb{E}(Z_k^2)) \right)^2 \right] \right\}^{1/2} (\mathbb{P}(\Omega_b^c))^{1/2}, \quad (4.34)$$

and we find

$$\begin{aligned} \frac{1}{2}\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{1}_{\Omega_b^c}) &\leq \left(\frac{3}{2}\|g\|^2 + \text{pen}_{th}(m) + C'\Delta^2 m_n^2 \|g\|^2 \right) \mathbb{P}(\Omega_b^c) \\ &\quad + \mathbb{E}^{1/2} \left[\left(\frac{1}{n\Delta} \sum_{k=1}^n (Z_k^2 - \mathbb{E}(Z_1^2)) \right)^2 \right] \mathbb{P}^{1/2}(\Omega_b^c) \\ &\quad + \mathbb{E}^{1/2} \left(\left(\frac{4}{\pi\Delta^2} \int_{-\pi m_n}^{\pi m_n} |\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2 du \right)^2 \right) \mathbb{P}^{1/2}(\Omega_b^c). \end{aligned}$$

Then we apply Proposition 4.2 with $l = 2$ and get for $p \geq 2$:

$$\mathbb{E} \left(\left| \frac{1}{n\Delta} \sum_{k=1}^n Z_k^2 - \mathbb{E}(Z_1^2) \right|^p \right) \leq C_p \left(\frac{1}{n\Delta} \right)^{p/2}.$$

Thus, by taking $p = 2$,

$$\mathbb{E}^{1/2} \left(\left(\frac{1}{n\Delta} \sum_{k=1}^n (Z_k^2 - \mathbb{E}(Z_k^2)) \right)^2 \right) \leq \frac{C}{\sqrt{n\Delta}}.$$

Applying (4.7) for $p = 2$ (see Proposition 4.1) gives

$$\mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^4) \leq \frac{C\Delta^2}{n^2}.$$

Thus

$$\begin{aligned} \mathbb{E} \left(\left(\frac{4}{\pi\Delta^2} \int_{-\pi m_n}^{\pi m_n} |\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^2 du \right)^2 \right) &\leq \frac{32\pi m_n}{\pi^2 \Delta^4} \int_{-\pi m_n}^{\pi m_n} \mathbb{E}(|\hat{\theta}_\Delta(u) - \theta_\Delta(u)|^4) du \\ &\leq C' \frac{m_n^2 \Delta^2}{\Delta^4 n^2} \leq C' \end{aligned}$$

as $m_n \leq n\Delta$. We obtain:

$$\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbf{1}_{\Omega_b^c}) \leq C(1 + n^2\Delta^4) \mathbb{P}(\Omega_b^c) + C'(1 + \frac{1}{\sqrt{n\Delta}}) \mathbb{P}^{1/2}(\Omega_b^c). \quad (4.35)$$

Lastly, it follows from the Markov inequality that

$$\begin{aligned} \mathbb{P}(\Omega_b^c) &\leq \frac{1}{b^p} \mathbb{E} \left(\left| \frac{(1/n\Delta) \sum_{k=1}^n Z_k^2}{\mathbb{E}(Z_1^2/\Delta)} - 1 \right|^p \right) \\ &\leq \frac{1}{(\mathbb{E}(Z_1^2/\Delta)b)^p} \mathbb{E} \left(\left| \frac{1}{n\Delta} \sum_{k=1}^n Z_k^2 - \mathbb{E}(Z_1^2/\Delta) \right|^p \right). \end{aligned}$$

We find that, if $\mathbb{E}(|Z_1|^{2p}) < +\infty$ and $p \geq 2$,

$$\mathbb{P}(\Omega_b^c) \leq \frac{C_p}{(\mathbb{E}(Z_1^2/\Delta)b)^p} \frac{1}{(n\Delta)^{p/2}}. \quad (4.36)$$

Therefore, using (4.35) and the above inequality, if we take $p = 4$ (i.e. $\mathbb{E}(Z_1^8) < \infty$), we get

$$\mathbb{E}(\|\hat{g}_{\hat{m}} - g\|^2 \mathbb{I}_{\Omega_b^c}) \leq C/(n\Delta).$$

This ends step 2 and the proof of Theorem 4.1. \square

4.2 Estimation on a compact set

In this paragraph, we intend to proceed without Fourier inversion and directly use the fact that

$$\frac{1}{n\Delta} \sum_{k=1}^n Z_k \delta_{Z_k} = \hat{\mu}_n \quad (4.37)$$

approximates the measure $\mu^{(1)}(dx) = g(x)dx$ (δ_z denotes the Dirac measure at z). We use the same contrast $\gamma_n(t)$ as previously with a different interpretation. Recall that, for any function t such that t^* is compactly supported,

$$\gamma_n(t) = \|t\|^2 - \frac{2}{2\pi} \langle \frac{\hat{\theta}_\Delta}{\Delta}, t^* \rangle.$$

As $\hat{\theta}_\Delta/\Delta$ is the Fourier Transform of $\hat{\mu}_n$ (see (4.5)), we now consider, with the same notation and for any compactly supported function t ,

$$\gamma_n(t) = \|t\|^2 - 2\langle \hat{\mu}_n, t \rangle = \|t\|^2 - \frac{2}{n\Delta} \sum_{k=1}^n Z_k t(Z_k).$$

More precisely, we fix a compact interval $A = [a, b] \subset \mathbb{R}$ and focus on the estimation of

$$g_A := g \mathbb{I}_A. \quad (4.38)$$

In other words, the estimation is performed in the “time domain” instead of previously, the “frequency domain”. We consider a family $(\Sigma_m, m \in \mathcal{M}_n)$ of finite dimensional linear subspaces of $\mathbb{L}^2(A)$: $\Sigma_m = \text{span}\{\phi_\lambda, \lambda \in \Lambda_m\}$ where $\text{card}(\Lambda_m) = D_m$ is the dimension of Σ_m . The set $\{\phi_\lambda, \lambda \in \Lambda_m\}$ denotes an orthonormal basis of Σ_m . We shall denote by $\|f\|_A^2 = \int_A f^2(u)du$ for any function f .

For $m \geq 1$, we define a collection $(\tilde{g}_m, m \in \mathcal{M}_n)$ of estimators of g_A by:

$$\tilde{g}_m = \arg \min_{t \in \Sigma_m} \gamma_n(t). \quad (4.39)$$

4.2.1 Projection spaces and their fundamental properties

We consider projection spaces classically used for density estimation on a compact set and satisfying the following conditions:

- (M1) $(\Sigma_m)_{m \in \mathcal{M}_n}$ is a collection of finite-dimensional linear sub-spaces of $\mathbb{L}^2(A)$, with dimension D_m such that $\forall m \in \mathcal{M}_n, D_m \leq n\Delta$. For all m , functions in Σ_m are of class C^1 in A , and, satisfy

$$\exists \Phi_0 > 0, \forall m \in \mathcal{M}_n, \forall t \in \Sigma_m, \|t\|_\infty \leq \Phi_0 \sqrt{D_m} \|t\|_A, \text{ and } \|t'\|_A \leq \Phi_0 D_m \|t\|_A. \quad (4.40)$$

where $\|t\|_\infty = \sup_{x \in A} |t(x)|$.

- (M2) $(\Sigma_m)_{m \in \mathcal{M}_n}$ is a collection of nested models, all embedded in a space \mathcal{S}_n belonging to the collection $(\forall m \in \mathcal{M}_n, \Sigma_m \subset \mathcal{S}_n)$. We denote by N_n the dimension of \mathcal{S}_n : $\dim(\mathcal{S}_n) = N_n$ ($\forall m \in \mathcal{M}_n, D_m \leq N_n \leq n\Delta$).

Inequality (4.40) is often referred to as the *norm connection* property of the projection spaces and is the basic tool to obtain the adequate order of the risk bound. This inequality should be compared with inequality 4.16 where the cut-off parameter plays the role of the dimension. It follows from Lemma 1 in [10], that (4.40) is equivalent to

$$\exists \Phi_0 > 0, \left\| \sum_{\lambda \in \Lambda_m} \phi_\lambda^2 \right\|_\infty \leq \Phi_0^2 D_m. \quad (4.41)$$

Functions of the spaces Σ_m are considered as functions on \mathbb{R} equal to zero outside A .

Here are the examples we have in view, and that we describe with $A = [0, 1]$ for simplicity. They satisfy assumptions (M1) and (M2).

[T] *Trigonometric spaces*, generated by $\phi_0 = 1_{[0,1]}$, $\phi_j(x) = \sqrt{2} \cos(2\pi jx) \mathbb{I}_{[0,1]}(x)$ and $\phi_{j+m+1}(x) = \sqrt{2} \sin(2\pi jx) \mathbb{I}_{[0,1]}(x)$ for $j = 1, \dots, m$, $D_m = 2m + 1$ and $\mathcal{M}_n = \{1, \dots, \lfloor n\Delta/2 \rfloor - 1\}$.

[W] *Dyadic wavelet generated spaces* with regularity $r \geq 2$ and compact support, as described e.g. in [38]. The generating basis is of cardinality $D_m = 2^{m+1}$ and $m \in \mathcal{M}_n = \{1, 2, \dots, \lfloor \log(n\Delta)/2 \rfloor - 1\}$.

4.2.2 Integrated risk on a compact set

Now, we have (see (4.39))

$$\tilde{g}_m = \sum_{\lambda \in \Lambda_m} \tilde{a}_\lambda \phi_\lambda \text{ with } \tilde{a}_\lambda = \frac{1}{n\Delta} \sum_{k=1}^n Z_k \phi_\lambda(Z_k). \quad (4.42)$$

And, for any $t \in \Sigma_m$,

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \tilde{g}_m \rangle = \|t - \tilde{g}_m\|^2 - \|\tilde{g}_m\|^2.$$

(for functions with support in A , $\|\cdot\| = \|\cdot\|_A$ and $\langle \cdot, \cdot \rangle_A = \langle \cdot, \cdot \rangle$). Let g_m denote the orthogonal projection of g_A on Σ_m , now given by

$$g_m = \sum_{\lambda \in \Lambda_m} a_\lambda \phi_\lambda \quad \text{with} \quad a_\lambda = \int_A t(x) g(x) dx = \langle t, g \rangle_A = \langle t, g \rangle.$$

At this stage, note that the “time domain approach” differs from the “frequency domain approach” only through the projection spaces. For simplicity, we use the same notation g_m to define the orthogonal projection of g_A on Σ_m . The contrast decomposition is the same

$$\gamma_n(t) - \gamma_n(s) = \|t - g\|^2 - \|s - g\|^2 - 2v_n(t - s) - 2R_n(t - s), \quad (4.43)$$

where the same v_n, R_n can be written now

$$v_n(t) = \frac{1}{n\Delta} \sum_{k=1}^n (Z_k t(Z_k) - e(Z_1 t(Z_1))), \quad (4.44)$$

and

$$R_n(t) = \frac{1}{\Delta} \mathbb{E}(Z_1 t(Z_1)) - \int t(x) g(x) dx. \quad (4.45)$$

This remainder term is ruled by the following proposition.

Proposition 4.5 *Let $t \in \Sigma_m$ and assume that (H1-g) and (H3-g) hold.*

1) *If $L := \int u^2 |g^*(u)|^2 du < +\infty$, then*

$$|R_n(t)| \leq \Delta \|t\|_A \|g\|_1 L^{1/2} / \sqrt{2\pi}.$$

2) *If g is bounded, $|R_n(t)| \leq C \Phi_0 \|t\|_A \Delta D_m$ where C depends on $\|g\|_1$, $\|g\|$, $\|g\|_\infty$ and A .*

3) *Otherwise:*

$$|R_n(t)| \leq C \Phi_0 \|t\|_A (\sqrt{\Delta D_m} + \Delta D_m), \quad (4.46)$$

where C depends on $\|g\|_1$, $\|g\|$ and A . If $n\Delta^2 \leq 1$, $|R_n(t)| = O(\sqrt{\Delta D_m})$.

Proof. First, we know that $R_n(t) = (1/2\pi) \int (\varphi_\Delta(u) - 1) g^*(u) t^*(-u) du$. Thus, if $\int u^2 |g^*(u)|^2 du < +\infty$, it follows from (3.11) that

$$R_n^2(t) \leq \frac{\Delta^2 \|g\|_1^2}{(2\pi)^2} \left(\int |u g^*(u) t^*(-u)| du \right)^2 \leq \frac{\Delta^2 \|g\|_1^2}{(2\pi)^2} \int u^2 |g^*(u)|^2 du \int |t^*(-u)|^2 du.$$

Noting that $\int |t^*(-u)|^2 du = 2\pi \|t\|^2 = 2\pi \|t\|_A^2$ gives 1).

For the two other cases, using Proposition 3.3, we have, for t a function with support $A = [a, b]$:

$$\frac{1}{\Delta} \mathbb{E}(Z_1 t(Z_1)) = \int_a^b t(z) \mathbb{E}g(z - Z_1) dz = \mathbb{E} \left(\int_{a-Z_1}^{b-Z_1} t(x + Z_1) g(x) dx \right).$$

Thus

$$R_n(t) = \mathbb{E} \left(\int_{a-Z_1}^{b-Z_1} t(x+Z_1)g(x)dx - \int_a^b t(x)g(x)dx \right).$$

On $(|Z_1| > b-a)$, $[a-Z_1, b-Z_1] \cap [a, b] = \emptyset$ and we use the bound

$$|R_n(t)| \leq 2\|t\|_\infty \|g\|_1.$$

We apply the Markov inequality, the norm connection (4.40) and the inequality $\mathbb{E}|Z_1| \leq \Delta \|g\|_1$ (see Proposition 3.2) to obtain:

$$\mathbb{E} \left(\mathbb{1}_{|Z_1| > b-a} |R_n(t)| \right) \leq 2\|t\|_\infty \|g\|_1 \frac{\mathbb{E}(|Z_1|)}{b-a} \leq \frac{2\Phi_0 \|g\|_1^2 \sqrt{D_m} \Delta \|t\|_A}{b-a}. \quad (4.47)$$

On $(|Z_1| \leq b-a)$, $[a-Z_1, b-Z_1] \cap [a, b] \neq \emptyset$. Assume for instance that $0 \leq Z_1 \leq b-a$. Then,

$$R_n(t) = \int_{a-Z_1}^a t(x+Z_1)g(x)dx + \int_a^{b-Z_1} (t(x+Z_1) - t(x))g(x)dx - \int_{b-Z_1}^b t(x)g(x)dx.$$

To study the middle term, we use the fact that t is C^1 on $[a, b]$.

$$\begin{aligned} T_1 &:= \mathbb{E} \left(\mathbb{1}_{0 \leq Z_1 \leq b-a} \int_a^{b-Z_1} (t(x+Z_1) - t(x))g(x)dx \right) \\ &= \mathbb{E} \left(Z_1 \mathbb{1}_{0 \leq Z_1 \leq b-a} \int_a^{b-Z_1} \int_0^1 t'(x+uZ_1)du g(x)dx \right) \\ &= \mathbb{E} \left(Z_1 \mathbb{1}_{0 \leq Z_1 \leq b-a} \int_0^1 \left(\int_a^{b-Z_1} t'(x+uZ_1)g(x)dx \right) du \right) \end{aligned}$$

An application of the Cauchy-Schwarz inequality yields

$$|T_1| \leq \mathbb{E}|Z_1| \|t'\|_A \|g\| \leq \Phi_0 \|g\|_1 \|g\| \|t\|_A \Delta D_m.$$

Next,

$$T_2 := \mathbb{E} \left(\mathbb{1}_{0 \leq Z_1 \leq b-a} \int_{a-Z_1}^a t(x+Z_1)g(x)dx \right).$$

Here we distinguish between 2) and 3). If g is bounded (case 2)), then, with $\mathbb{E}(|Z_1|) \leq \Delta \|g\|_1$ and (4.40), we obtain:

$$|T_2| \leq \|t\|_\infty \|g\|_\infty \mathbb{E}(|Z_1|) \leq \Phi_0 \|g\|_\infty \|g\|_1 \|t\|_A \Delta \sqrt{D_m}.$$

Otherwise (case 3)), using the Cauchy-Schwarz inequality again,

$$\begin{aligned} |T_2| &\leq \mathbb{E}(\sqrt{Z_1^+}) \|t\|_\infty \|g\| \leq \sqrt{\mathbb{E}(|Z_1|)} \Phi_0 \sqrt{D_m} \|t\|_A \|g\| \\ &\leq \Phi_0 \|t\|_A \sqrt{\|g\|_1} \|g\| \sqrt{D_m} \Delta. \end{aligned}$$

The same bound holds for the last term.

The same study can be done for $a - b \leq Z_1 \leq 0$. Joining all terms, we find that, if g is bounded

$$|R_n(t)| \leq C\Phi_0 \|t\|_A \Delta D_m.$$

Otherwise,

$$|R_n(t)| \leq C'\Phi_0 \|t\|_A (\sqrt{\Delta D_m} + \Delta D_m).$$

The constants C and C' depend on $a, b, \|g\|_1$ and $\|g\|$. Recalling that $D_m \leq n\Delta$, we have, as $n\Delta^2 \leq 1$, that $|R_n(t)| = O(\sqrt{\Delta D_m})$.

Proposition 4.6 *Assume that (H1-g)-(H2)(2)-(H3-g) hold. The estimator \tilde{g}_m of g_A (see (4.39)) satisfies*

$$\mathbb{E}(\|\tilde{g}_m - g\|_A^2) \leq 3\|g - g_m\|_A^2 + 16\Phi_0[\mathbb{E}(Z_1^2)/\Delta] \frac{D_m}{n\Delta} + K\rho_{m,\Delta}, \quad (4.48)$$

where g_m is the orthogonal projection of g_A on Σ_m . The constant K depends on m_1, m_2 (see Proposition 3.1) and g . The remainder term satisfies $\rho_{m,\Delta} = \Delta^2$ if $\int u^2 |g^*(u)|^2 du < +\infty$, $\rho_{m,\Delta} = \Delta^2 D_m^2$ if g is bounded. Otherwise $\rho_{m,\Delta} = \Delta D_m$ if $n\Delta^2 \leq 1$.

Proof. Relation (4.25) still holds with v_n and R_n respectively defined by (4.44) and (4.45). As for any $t \in \Sigma_m$,

$$\|t - g\|^2 = \|t - g\|_A^2 + \|g\|_{A^c}^2,$$

we get

$$\gamma_n(t) - \gamma_n(s) = \|t - g\|_A^2 - \|s - g\|_A^2 - 2v_n(t - s) - 2R_n(t - s).$$

Writing that $\gamma_n(\tilde{g}_m) - \gamma_n(g) \leq \gamma_n(g_m) - \gamma_n(g)$, we get

$$\|\tilde{g}_m - g\|_A^2 \leq \|g_m - g\|_A^2 + 2v_n(\tilde{g}_m - g_m) + 2R_n(\tilde{g}_m - g_m).$$

We have

$$2v_n(\tilde{g}_m - g_m) \leq \frac{1}{8}\|\tilde{g}_m - g_m\|_A^2 + 8 \sup_{t \in \Sigma_m, \|t\|_A=1} [v_n(t)]^2,$$

and the analogous inequality for R_n . Using that

$$\|\tilde{g}_m - g_m\|_A^2 \leq 2\|g - g_m\|_A^2 + 2\|\tilde{g}_m - g\|_A^2$$

and some algebra yields:

$$\frac{1}{2}\|\tilde{g}_m - g\|_A^2 \leq \frac{3}{2}\|g_m - g\|_A^2 + 8 \sup_{t \in \Sigma_m, \|t\|_A=1} [v_n(t)]^2 + 8 \sup_{t \in \Sigma_m, \|t\|_A=1} [R_n(t)]^2.$$

To bound the last term, we use Proposition 4.5. Noting that each $t \in \Sigma_m$ can be written $t = \sum_{\lambda \in \Lambda_m} t_\lambda \phi_\lambda$ with $\sum t_\lambda^2 = 1$ if $\|t\|_A = 1$, we get

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \Sigma_m, \|t\|_A=1} [v_n(t)]^2 \right) &\leq \sum_{\lambda \in \Lambda_m} \mathbb{E}([v_n(\phi_\lambda)]^2) = \sum_{\lambda \in \Lambda_m} \frac{1}{n\Delta^2} \text{Var}(Z_1 \phi_\lambda(Z_1)) \\ &\leq \mathbb{E}(Z_1^2 \sum_{\lambda} \phi_\lambda^2(Z_1)) \frac{1}{n\Delta^2} = [\mathbb{E}(Z_1^2)/\Delta] \frac{\Phi_0 D_m}{n\Delta}. \end{aligned} \quad (4.49)$$

We have used (4.41) in the last line. The conclusion of Proposition 4.6 follows .

As for Proposition 4.3, we draw the consequences of Proposition 4.6 on the rate of convergence of the risk bound. In the setting of this section, the regularity of g_A must be described by using classical Besov spaces on compact sets. Let us recall that the Besov space $\mathcal{B}_{\alpha,2,\infty}([0,1])$ is defined by:

$$\mathcal{B}_{\alpha,2,\infty}([0,1]) = \{f \in \mathbb{L}^2([0,1]), |f|_{\alpha,2} := \sup_{t>0} t^{-\alpha} \omega_r(f,t)_2 < +\infty\}$$

where $r = [\alpha] + 1$ ($[\cdot]$ denotes the integer part), and $\omega_r(f,t)_2$ is the r -th modulus of smoothness of a function $f \in \mathbb{L}^2([0,1])$ and is equal to:

$$\omega_r(f,t)_2 = \sup_{0 < h \leq t} \|\Delta_h^r(f, \cdot)\|_2([0,1-rh]), t \geq 0, \Delta_h^r(f,x) = \sum_{k=0}^r \binom{r}{k} (-1)^{r-k} f(x+kh).$$

Note that $|f|_{\alpha,2}$ is a semi-norm with usual associated norm $\|f\|_{\alpha,2} = \|f\| + |f|_{\alpha,2}$. For details, we refer to [23], p.54-57.

Heuristically, a function in $\mathcal{B}_{\alpha,2,\infty}([0,1])$ can be seen as square integrable and $[\alpha]$ -times differentiable with derivative of order $[\alpha]$ having a Hölder property of order $\alpha - [\alpha]$.

Proposition 4.7 Consider $A = [0,1]$ and Σ_m a space in collection [T] or [W]. Assume that (H1-g), (H2)(2) and (H3-g) hold. Let $g_{[0,1]} \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, $D_m = (n\Delta)^{1/(2\alpha+1)}$. Assume that we can choose Δ of the form $\Delta = n^{-a}$ with $a \in (0,1)$ and:

- $a \geq \alpha/(3\alpha+1)$, if $\int u^2 |g^*(u)|^2 du < +\infty$,
- $a \geq (\alpha+1)/(3\alpha+2)$, if g is bounded,
- $a \geq 1/2$, otherwise.

Then

$$\mathbb{E}(\|g - \tilde{g}_m\|_A^2) \leq K(n\Delta)^{-2\alpha/(2\alpha+1)}.$$

Proof. In [23], it is proved that, if Σ_m is a space of [T] or [W], and if $g \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, then

$$\|g - g_m\|_{[0,1]}^2 \leq CD_m^{-2\alpha}.$$

Minimizing $D_m^{-2\alpha} + D_m/(n\Delta)$ leads to the best choice $D_m = O((n\Delta)^{1/(2\alpha+1)})$ for which the first two terms in (4.48) have the same rate $O((n\Delta)^{-2\alpha/(2\alpha+1)})$.

Now, we search for the choice of $\Delta = n^{-a}$ such that the remainder term satisfies

$$\rho_{m,\Delta} \leq (n\Delta)^{-2\alpha/(2\alpha+1)}.$$

We distinguish the cases of Proposition 4.5.

If $\int u^2 |g^*(u)|^2 du < +\infty$, $\rho_{m,\Delta} = \Delta^2$ and we find $a \geq \alpha/(3\alpha + 1)$.

If g is bounded, $\rho_{m,\Delta} = \Delta^2 D_m^2$ and we find $a \geq (\alpha + 1)/(3\alpha + 2)$. Otherwise, $\rho_{m,\Delta} = \Delta D_m$ and we find $a \geq 1/2$.

Note that $a \geq \alpha/(3\alpha + 1)$ and $a \geq (\alpha + 1)/(3\alpha + 2)$ holds for any $\alpha \geq 0$ if $a \geq 1/3$ (hence $n\Delta \leq n^{2/3}$), and $a \geq 1/2$ implies $n\Delta \leq n^{1/2}$.

4.2.3 Adaptive result

Now, to get an adaptive result, we need to define a penalty function $\text{pen}(\cdot)$ and set

$$\tilde{m} = \arg \min_{m \in \mathcal{M}_n} (\gamma_n(\tilde{g}_m) + \text{pen}(m))$$

Let

$$\text{pen}(m) = \frac{\kappa}{n\Delta} \sum_{k=1}^n Z_k^2 \frac{D_m}{n\Delta}, \quad \text{pen}_{th}(m) = \mathbb{E}(\text{pen}(m)) = \kappa \mathbb{E}(Z_1^2 / \Delta) \frac{D_m}{n\Delta}.$$

Here too, we use the same notation $\text{pen}(m)$, $\text{pen}_{th}(m)$ as above, although the definitions differ. The following theorem holds:

Theorem 4.2 *Assume that assumptions (H1-g)-(H2)(12)-(H3-g) and conditions (M1)-(M2) for the collection of spaces are fulfilled. There exists a universal constant κ such that*

$$\mathbb{E}(\|g - \tilde{g}_{\tilde{m}}\|_A^2) \leq C \inf_{m \in \mathcal{M}_n} (\|g - g_m\|_A^2 + \text{pen}_{th}(m)) + C\rho_{n,\Delta} + \frac{C'}{n\Delta},$$

where $\rho_{n,\Delta} = \Delta^2$ if $\int u^2 |g^*(u)|^2 du < +\infty$, $\rho_{n,\Delta} = \Delta^2 N_n^2$ if g is bounded. Otherwise, $\rho_{n,\Delta} = \Delta N_n$.

Remark 4.3 *The moment condition of order 12 in Theorem 4.2 can be weakened into a condition of order 8 for basis [T], which is bounded.*

A subsection below is devoted to the proof of Theorem 4.2. We deduce the following corollary.

Corollary 4.1 *Let the Σ_m 's be D_m -dimensional linear spaces in collections [T] or [W]. Assume moreover that g belongs to $\mathcal{B}_{\alpha,2,\infty}([0,1])$ with $r > \alpha > 0$ and that we can choose $\Delta = n^{-a}$ with $a \in [1/3, 1[$ if $\int u^2 |g^*(u)|^2 du < +\infty$, $a \in [3/5, 1[$ if g is bounded, and otherwise, $a \in [2/3, 1[$. Then, under the assumptions of Theorem 4.2,*

$$\mathbb{E}(\|g - \tilde{g}_{\tilde{m}}\|^2) = O\left((n\Delta)^{-\frac{2\alpha}{2\alpha+1}}\right). \quad (4.50)$$

Remark 4.4 *The bound α on r stands for the regularity of the basis functions for collection [W]. For the trigonometric collection [T], no such bound is required.*

Proof. We apply results of [23] and Lemma 12 of [6]. If $g \in \mathcal{B}_{\alpha,2,\infty}([0,1])$ for some $\alpha > 0$, then $\|g - g_m\|$ is of order $D_m^{-\alpha}$ in the collections [T] and [W]. Thus the infimum in Theorem 4.2 is reached for $D_{m_n} = O([(n\Delta)^{1/(1+2\alpha)}])$, which is less than $n\Delta$ for $\alpha > 0$.

Now, we look at the remainder term and find conditions on Δ implying that

$$\rho_{n,\Delta} \leq (n\Delta)^{-1}.$$

Recall that the maximal dimension N_n of the models collection satisfies $N_n \leq n\Delta$.

If $\int u^2 |g^*(u)|^2 du < +\infty$, $\Delta^2 \leq 1/(n\Delta)$ holds for $\Delta = n^{-a}$ if $a \in [1/3, 1[$.

If g is bounded, $\Delta^2 N_n^2 \leq 1/(n\Delta)$ holds if $\Delta^2 (n\Delta)^2 \leq 1/(n\Delta)$ which gives $a \in [3/5, 1[$.

Otherwise, $N_n \Delta \leq 1/(n\Delta)$ holds for $\Delta = n^{-a}$ if $a \in [2/3, 1[$. Unfortunately, this also implies that $n\Delta \leq n^{2/3}$ in the first case, $n\Delta \leq n^{2/5}$ in the second case and $n\Delta \leq n^{1/3}$ in the third case. Then, we find the standard nonparametric rate of convergence $(n\Delta)^{-2\alpha/(1+2\alpha)}$.

Remark 4.5 In [33], the nonparametric estimation of $n(\cdot)$ from a continuous observation $(L_t)_{t \in [0,T]}$ is investigated. The authors use projection methods and penalization to obtain estimators with rate $O(T^{-2\alpha/(2\alpha+1)})$ on a Besov class $\mathcal{B}_{\alpha,2,\infty}([0,1])$. Moreover, in [31], a minimax bound for the estimation of $n(\cdot)$ based on discrete observations of order $O((n\Delta)^{-2\alpha/(2\alpha+1)})$ is obtained. The results can therefore be compared since rates are identical, except that we do not estimate the same function.

4.2.4 Proof of Theorem 4.2

The proof of Theorem 4.2 is close to the proof of Theorem 4.1. Hence we focus mainly on the differences. Note that v_n defined in (4.44) can be written as

$$v_n(t) = \frac{1}{n} \sum_{k=1}^n (f_t(Z_k) - \mathbb{E}(f_t(Z_1)))$$

with f_t now given by $f_t(z) = zt(z) = z \mathbb{I}_{z \in A} t(z)$, since t has compact support A . As in step 1 of Theorem 4.1, we are led to the inequality:

$$\begin{aligned} \frac{1}{2} \|\tilde{g}_m - g\|_A^2 \mathbb{I}_{\Omega_b} &\leq \frac{3}{2} \|g - g_m\|_A^2 + 2\text{pen}_{t_h}(m) \\ &+ 8 \sum_{m' \in \mathcal{M}_n} \left(\sup_{t \in \Sigma_m + \Sigma_{m'}, \|t\|_A=1} [v_n(t)]^2 - p(m, m') \right) + \\ &+ 8 \sup_{t \in \mathcal{S}_n, \|t\|_A=1} [R_n(t)]^2, \end{aligned}$$

with $8p(m, m') \leq (1-b)(\text{pen}_{t_h}(m) + \text{pen}_{t_h}(m'))$, for all $m \in \mathcal{M}_n$.

It follows from Proposition (4.5) that

$$\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [R_n(t)]^2 \leq K\rho_{n,\Delta}.$$

The function $p(m, m')$ is chosen in order to ensure the following Lemma.

Lemma 4.2 *Under the Assumptions of Theorem 4.2, define*

$$p(m, m') = 4\mathbb{E}(Z_1^2/\Delta) \frac{D_m \vee D_{m'}}{n\Delta}, \quad (4.51)$$

then

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left(\sup_{t \in \Sigma_m + \Sigma_{m'}, \|t\|_A=1} [v_n(t)]^2 - p(m, m') \right)_+ \leq \frac{C}{n\Delta},$$

where C is a constant.

For the study of $\mathbb{E}(\|\tilde{g}_{\hat{m}} - g\|_A^2 \mathbb{I}_{\Omega_b^c})$, as in step 2 above, we have the inequality analogous to (4.32):

$$\frac{1}{2} \|\hat{g}_{\hat{m}} - g\|^2 \leq \frac{3}{2} \|g_A - g_m\|^2 + \text{pen}(m) + 8 \sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^2 + 8 \sup_{t \in \mathcal{S}_n, \|t\|_A=1} [R_n(t)]^2.$$

The bound for $\mathbb{P}(\Omega_b^c)$ is given by (4.36). Proposition 4.5 applies to bound $[R_n(t)]^2$ by $C\rho_{n,\Delta}$.

Then we have again

$$\text{pen}(m) \mathbb{I}_{\Omega_b^c} \leq \text{pen}_{t_h}(m) + (\text{pen}(m) - \text{pen}_{t_h}(m)) \mathbb{I}_{\Omega_b^c}.$$

The same bound holds also for the term $\mathbb{E}[(\text{pen}(m) - \mathbb{E}(\text{pen}(m))) \mathbb{I}_{\Omega_b^c}]$. We apply inequality (4.34).

It remains to study the term $\mathbb{E}(\sup_{t \in \mathcal{S}_n} [v_n(t)]^2 \mathbb{I}_{\Omega_b^c})$. We use

$$\mathbb{E} \left(\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^2 \mathbb{I}_{\Omega_b^c} \right) \leq \left(\mathbb{E} \sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^4 \right)^{1/2} \mathbb{P}^{1/2}(\Omega_b^c).$$

Denote by $(\phi_\lambda)_{\lambda \in \Lambda_n}$ an orthonormal basis of \mathcal{S}_n , $|\Lambda_n| = N_n$. We have

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^4 \right) &= \mathbb{E} \left[\left(\sum_{\lambda \in \Lambda_n} v_n^2(\phi_\lambda) \right)^2 \right] \\ &\leq N_n \sum_{\lambda \in \Lambda_n} \mathbb{E} \left\{ \left(\frac{1}{n\Delta} \sum_{k=1}^n (Z_k \phi_\lambda(Z_k) - \mathbb{E}(Z_k \phi_\lambda(Z_k))) \right)^4 \right\} \\ &\leq \frac{KN_n}{(n\Delta)^4} \sum_{\lambda \in \Lambda_n} \left[n\mathbb{E}[(Z_1 \phi_\lambda(Z_1))^4] + (n\mathbb{E}(Z_1^2 \phi_\lambda^2(Z_1)))^2 \right], \end{aligned}$$

where the last inequality follows from the Rosenthal Inequality (.1).

If the basis is bounded, $\phi_\lambda^2 \leq B$, $\forall \lambda$, as for instance basis [T] ($B = 2$), we find

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^4 \right) &\leq \frac{KN_n^2 B^2}{(n\Delta)^4} [n\mathbb{E}(Z_1^4/\Delta)\Delta + n^2\mathbb{E}^2(Z_1^2/\Delta)\Delta^2] \\ &\leq \frac{K'N_n^2}{(n\Delta)^2} \leq K' \end{aligned}$$

using $N_n \leq n\Delta$.

In the general case, we use that $\sum_\lambda \phi_\lambda^4(x) \leq \|\phi_\lambda\|_\infty^2 \sum_\lambda \phi_\lambda^2(x)$ and $\|\sum_\lambda \phi_\lambda^2\|_\infty \leq \Phi_0^2 N_n$ and $\|\phi_\lambda\|_\infty^2 \leq \Phi_0^2 N_n$, so that

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^4 \right) &\leq \frac{KN_n}{(n\Delta)^4} \left[\Phi_0^4 N_n^2 n \mathbb{E}(Z_1^4/\Delta)\Delta + n^2 \mathbb{E}^2 \left(\sum_{\lambda \in \Lambda_n} (Z_1^2/\Delta) \phi_\lambda^2(Z_1) \right) \Delta^2 \right] \\ &\leq \frac{KN_n}{(n\Delta)^4} \left[\Phi_0^4 N_n^2 n \mathbb{E}(Z_1^4/\Delta)\Delta + n^2 \Phi_0^4 N_n^2 \mathbb{E}^2(Z_1^2/\Delta)\Delta^2 \right] \\ &\leq \frac{K''N_n^3}{(n\Delta)^2} \leq K''(n\Delta) \end{aligned}$$

using $N_n \leq n\Delta$.

Using (4.36), we obtain $\mathbb{E} \left(\sup_{t \in \mathcal{S}_n, \|t\|_A=1} [v_n(t)]^2 \mathbb{1}_{\Omega_b^c} \right) \leq C/(n\Delta)$ if $\mathbb{P}(\Omega_b^c) \leq 1/(n\Delta)^2$ which holds for $p = 4$ and $\mathbb{E}(Z_1^8) < +\infty$ in the first case (bounded basis). In the general case, we need $\mathbb{P}(\Omega_b^c) \leq 1/(n\Delta)^3$ and thus $p = 6$ and $\mathbb{E}(Z_1^{12}) < +\infty$.

4.2.5 Proof of Lemma 4.2

Again, we apply the Talagrand (see Appendix) Inequality to the class

$$\mathcal{F} = \{f_t, t \in \Sigma_m + \Sigma_{m'}\} \text{ where } f_t(z) = \frac{z \mathbb{1}_{z \in At}(z)}{\Delta}.$$

We obtain similarly to (4.49)

$$H^2 = [\mathbb{E}(Z_1^2)/\Delta] \Phi_0(D_m \vee D_{m'})/(n\Delta) \text{ and } M = b_A \Phi_0 \sqrt{D_m \vee D_{m'}}/\Delta,$$

where $b_A = \sup_{z \in A} |z|$. Lastly, we find

$$\begin{aligned}
\text{Var}\left(\frac{Z_1}{\Delta}t(Z_1)\right) &\leq \mathbb{E}(Z_1^2 t^2(Z_1))/\Delta^2 = \frac{1}{\Delta} \int z t^2(z) \mathbb{E}(g(z - Z_1)) dz \\
&\leq \frac{b_A \|t\|_\infty}{\Delta} \mathbb{E}\left(\int |t(z)g(z - Z_1)| dz\right) \\
&\leq \frac{b_A \Phi_0 (D_m \vee D_{m'})^{1/2}}{\Delta} \mathbb{E}\left(\|t\| \int g^2(z - Z_1) dz\right)^{1/2} \\
&\leq \frac{2b_A \Phi_0 (D_m \vee D_{m'})^{1/2} \|g\|}{\Delta}.
\end{aligned}$$

We denote by $v = C(D_m \vee D_{m'})^{1/2}/\Delta$ with $C = 2\Phi_0 b_A \|g\|$.

Then we get

$$\begin{aligned}
\mathbb{E}\left(\sup_{t \in \Sigma_m + \Sigma_{m'}, \|t\|_A=1} [v_n(t)]^2 - p(m, m')\right)_+ &\leq C'_1 \left(\frac{\sqrt{D_m \vee D_{m'}}}{n\Delta} e^{-C_2 \sqrt{D_m \vee D_{m'}}}\right. \\
&\quad \left. + \frac{1}{n\Delta} \exp(-\sqrt{n\Delta})\right).
\end{aligned}$$

Therefore, as $D_m \leq n\Delta$, as above

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E}\left(\sup_{t \in \Sigma_m + \Sigma_{m'}, \|t\|_A=1} [v_n(t)]^2 - p(m, m')\right)_+ \leq \frac{C}{n\Delta}.$$

This ends the proof of Lemma 4.2. \square

4.3 Kernel estimators

The fact that $(1/(n\Delta)) \sum_{k=1}^n Z_k \delta_{Z_k} = \hat{\mu}_n$ approximates the measure $\mu^{(1)}(dx) = g(x)dx$ can be used to build kernel estimators of g . Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a kernel, *i.e.* an integrable function such that

$$\int K(u) du = 1. \tag{4.52}$$

As it is usual, we assume that K is an even function. Set $K_h(x) = \frac{1}{h} K(\frac{x}{h})$ and define the kernel estimator of g with bandwidth h by:

$$\hat{g}_h(x) = K_h \star \hat{\mu}_n(x) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k K_h(x - Z_k). \tag{4.53}$$

The kernel estimator (4.53) can be related to the deconvolution estimator (4.10). Indeed, let us compute the Fourier transform of \hat{g}_h :

$$(\hat{g}_h)^*(u) = \frac{1}{n\Delta h} \sum_{k=1}^n Z_k \int K\left(\frac{x-Z_k}{h}\right) e^{iux} dx.$$

After a change of variable, we obtain (see (4.10)):

$$(\hat{g}_h)^*(u) = \frac{1}{n\Delta} \sum_{k=1}^n Z_k e^{iuZ_k} K^*(uh) = \frac{\hat{\theta}_\Delta(u)}{\Delta} K^*(uh).$$

Under the assumption that K^* is integrable, we have:

$$\hat{g}_h(x) = \frac{1}{2\pi} \int e^{-ixu} \frac{\hat{\theta}_\Delta(u)}{\Delta} K^*(uh) du. \quad (4.54)$$

Thus, the kernel estimator \hat{g}_h is obtained as the deconvolution estimator (4.10) using another kernel than ϕ (see (4.12)) and with the correspondence $h = m^{-1}$. Moreover, the inequality

$$|\hat{g}_h(x)| \leq \frac{1}{2\pi\Delta} \sum_{k=1}^n |Z_k| \int |K^*(uh)| du$$

implies that $\hat{g}_h(x)$ is integrable as $\mathbb{E}|Z_k| < +\infty$ by (H1-g).

4.3.1 Mean integrated squared error for fixed bandwidth

To study the MISE of the kernel estimator \hat{g}_h , we precise assumptions on the kernel K and additional assumptions on g . For $\alpha > 0$, we denote by $l = \lfloor \alpha \rfloor$ the largest integer strictly smaller than α . The following definition is classical.

Definition 4.1 A kernel K is said to be of order l if functions $u \mapsto u^j K(u)$, $j = 0, 1, \dots, l$ are integrable and satisfy

$$\int u^j K(u) du = 0, \forall j \in \{1, \dots, l\}. \quad (4.55)$$

The assumptions on K are the following.

- (Ker[1]) For some $\alpha > 0$, K is a kernel of order $l = \lfloor \alpha \rfloor$ and $\int |x|^\alpha |K(x)| dx < +\infty$.
- (Ker[2]) $\|K\|_2 < +\infty$.
- (Ker[3]) $K^* \in \mathbb{L}^1$.

Assumptions (Ker[i]), $i = 1, 2$ are standard assumptions when working on problems of estimation by kernel methods. As noted above, (Ker[3]) is more specific and ensures in particular that $\hat{g}_h(x)$ is integrable under (H1-g).

Remark 4.6 To construct a kernel of order l , we may proceed as follows. Choose u an even and integrable function such that $u \in \mathbb{L}^2(\mathbb{R})$, $u^* \in \mathbb{L}^1(\mathbb{R})$, $\int u(y) dy = 1$ and $\int |y|^k |u(y)| dy < +\infty$, and define for any given integer l ,

$$K(t) = \sum_{k=1}^{l+1} \binom{l+1}{k} (-1)^{k+1} \frac{1}{k} u \left(\frac{t}{k} \right) \quad (4.56)$$

The kernel K defined by (4.56) is order l and satisfies $(Ker[i])$ $i = 1, 2, 3$ (see [48] and [34]).

The definition of kernels of order l satisfying $(Ker[1])$ is fitted to evaluate the bias of kernel estimators on Nikol'ski classes of functions.

Definition 4.2 (*Nikol'ski class*) Let $\alpha > 0$, $L > 0$. Let also $l = \lfloor \alpha \rfloor$ be the largest integer strictly smaller than α . The Nikol'ski class $N(\alpha, L)$ on \mathbb{R} is the set of all functions $f : \mathbb{R} \rightarrow \mathbb{R}$ such that derivatives $f^{(j)}$ for $j = 1, \dots, l$ exist and $f^{(l)}$ verifies:

$$\left(\int |f^{(l)}(x+t) - f^{(l)}(x)|^2 dx \right)^{1/2} \leq L|t|^{\alpha-l}, \forall t \in \mathbb{R}. \quad (4.57)$$

In addition to (H1-g), (H3-g) and some moment assumption (H2-(k)), we may require that g belongs to $N(\alpha, L)$.

The MISE of \hat{g}_h can be split using the standard bias variance decomposition:

$$\mathbb{E}[\|\hat{g}_h - g\|^2] = \int \mathbb{E}[(\hat{g}_h(x) - \mathbb{E}[\hat{g}_h(x)])^2] dx + \int (\mathbb{E}[\hat{g}_h(x)] - g(x))^2 dx$$

The bias needs further decomposition:

$$\begin{aligned} \mathbb{E}[\|\hat{g}_h - g\|^2] &\leq \int \text{Var}(\hat{g}_h(x)) dx + 2 \int (K_h \star g(x) - g(x))^2 dx \\ &\quad + 2 \int (\mathbb{E}[\hat{g}_h(x)] - K_h \star g(x))^2 dx \\ &:= \int \text{Var}(\hat{g}_h(x)) dx + 2 \int b_{h,1}^2(x) dx + 2 \int b_{h,2}^2(x) dx \end{aligned}$$

with the usual bias of the kernel method,

$$b_{h,1}(x) = K_h \star g(x) - g(x), \quad (4.58)$$

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

$$b_{h,2}(x) = \mathbb{E}[\hat{g}_h(x)] - K_h \star g(x). \quad (4.59)$$

In other words

$$b_h(x) = \mathbb{E}[\hat{g}_h(x)] - g(x) = b_{h,1}(x) + b_{h,2}(x). \quad (4.60)$$

The bias terms are bounded as follows.

Lemma 4.3 Under $(Ker[1])$ and if $g \in N(\alpha, L)$,

$$\|K_h \star g - g\|^2 = \|b_{h,1}\|^2 \leq c_1 h^{2\alpha}$$

with $c_1 = (L/l! \int |K(v)||v|^\alpha dv)^2$.

Assume (Ker[3]), (H1-g), (H3-g) and $\int u^2 |g^*(u)|^2 du := A < +\infty$. Then,

$$\|b_{h,2}\|^2 \leq c'_1 \Delta^2$$

with $c'_1 = A \|K\|_1^2 \|g\|_1^2 / 2\pi$.

Proof. Assumption (Ker[1]) and the fact that $g \in N(\alpha, L)$ standardly imply the inequality (see [61])

$$\int b_{h,1}^2(x) dx \leq c_1 h^{2\alpha}.$$

Thus, we focus on $b_{h,2}$. Under (Ker[3]), by the Fourier inversion formula, we have for all z ,

$$K\left(\frac{x-z}{h}\right) = \frac{1}{2\pi} \int e^{-i\frac{x-z}{h}v} K^*(v) dv = \frac{h}{2\pi} \int e^{iuz} e^{-iux} K^*(uh) du.$$

This shows that $|K(\frac{x-z}{h})|$ is bounded. Assumption (H1-g) ensures that $\mathbb{E}|Z_1| \leq \int |g(z)| dz < +\infty$. Thus, (see (4.3))

$$\begin{aligned} b_{h,2}(x) &= \frac{1}{h\Delta} \mathbb{E} \left[Z_1 K\left(\frac{x-Z_1}{h}\right) \right] - \frac{1}{h} \int K\left(\frac{x-z}{h}\right) g(z) dz \\ &= \frac{1}{2\pi} \int e^{-ixu} K^*(uh) \left(\frac{\theta_\Delta(u)}{\Delta} - g^*(u) \right) du. \end{aligned}$$

Therefore, we get, with the Parseval Formula and (4.4),

$$\|b_{h,2}\|^2 = \int b_{h,2}^2(x) dx = \frac{1}{2\pi} \int |K^*(uh)|^2 |\varphi_\Delta(u) - 1|^2 |g^*(u)|^2 du.$$

Now, applying Inequality (3.11) of Proposition 3.4, we get

$$\|b_{h,2}\|^2 \leq \frac{\|g\|_1^2 \Delta^2}{2\pi} \int |K^*(uh)|^2 u^2 |g^*(u)|^2 du.$$

Since $|K^*(uh)| \leq \|K\|_1 < +\infty$, we obtain the announced bound.

Moreover, the variance is controlled as follows:

Lemma 4.4 Under (Ker[2]), (Ker[3]), (H1-g), (H2-(2)) and (H3-g), we have

$$\int \text{Var}[\hat{g}_h(x)] dx \leq \frac{\|K\|^2 \mathbb{E}(Z_1^2/\Delta)}{nh\Delta}.$$

Proof. As the Z_k are i.i.d., we have:

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

Thus,

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

With the Fubini-Tonelli theorem, we get

$$\int \text{Var}[\hat{g}_h(x)] dx \leq \frac{1}{n(h\Delta)^2} \mathbb{E} \left[Z_1^2 \int K^2 \left(\frac{Z_1 - x}{h} \right) dx \right] = \frac{\|K\|^2 \mathbb{E}(Z_1^2)}{nh\Delta^2}.$$

This ends the proof of Lemma 4.4.

Recall that $\mathbb{E}(Z_1^2)/\Delta = m_2 + \Delta m_1^2$ by Proposition 3.1. Lemmas 4.3 et 4.4 lead us to the following risk bound.

Proposition 4.8 *Under (Ker[1]) to (Ker[3]), (H1-g), (H2-(2)), (H3-g) and if $\int v^2 |g^*(v)|^2 dv := A < +\infty$, we have*

$$\mathbb{E}(\|\hat{g}_h - g\|^2) \leq 2\|g - g \star K_h\|^2 + \frac{\|K\|^2 \mathbb{E}(Z_1^2/\Delta)}{nh\Delta} + (A\|K\|_1^2 \|g\|_1^2 / \pi) \Delta^2. \quad (4.61)$$

If in addition $g \in N(\alpha, L)$, then $\|g - g \star K_h\|^2 \leq c_1 h^{2\alpha}$ with c_1 given in Lemma 4.3.

4.3.2 Rates of convergence

We set $h = h_n$ with $h_n \rightarrow 0$ and $nh_n \rightarrow +\infty$. Recall that $\Delta = \Delta_n$ is such that $\lim_{n \rightarrow +\infty} \Delta_n = 0$. Consequently, $1/nh$ is negligible compared to $1/nh\Delta$. To obtain the optimal convergence rate based on the first two terms of (4.61), a constraint on Δ is necessary. We impose $\Delta^2 \leq 1/(nh\Delta)$, equivalently

$$\Delta^3 \leq \frac{1}{nh}. \quad (4.62)$$

The optimal choice of h_n is $h_{opt} \propto ((n\Delta)^{-\frac{1}{2\alpha+1}})$ and the associated rate has order $O\left((n\Delta)^{-\frac{2\alpha}{2\alpha+1}}\right)$. Therefore, we can state:

Proposition 4.9 *Under the assumptions of Proposition 4.8 and under condition (4.62) the choice $h_{opt} \propto ((n\Delta)^{-\frac{1}{2\alpha+1}})$ minimizes the risk bound (4.61) and gives*

$$\|\hat{g}_{h_{opt}} - g\|^2 = O((n\Delta)^{-\frac{2\alpha}{2\alpha+1}}).$$

4.3.3 Data-driven choice of the bandwidth and adaptive estimator

Now, α being unknown, we must select the bandwidth by a data-driven criterion. For this, adequate estimators of the dominating risk bound terms (see (4.61)) must be found. Following ideas given in [34] for density estimation, we set:

$$V(h) = \kappa \|K\|_1^2 \|K\|^2 \frac{\mathbb{E}(Z_1^2/\Delta)}{nh\Delta}, \quad (4.63)$$

where κ is a numerical constant that will be precised below. Note that $V(h)$ is proportional to the bound of $\int \text{Var}[\hat{g}_h(x)]dx$. In the above definition, $V(h)$ depends on the unknown moment $\mathbb{E}Z_1^2$. Actually, this moment should be replaced by the empirical mean $n^{-1} \sum_{k=1}^n Z_k^2$. This substitution is possible and can be done as in the proof of Theorem 4.1 by introducing the set Ω_b (see (4.21)) and splitting the proof into the analogous steps 1 and 2. For the sake of simplicity, we omit the substitution and only deal with the deterministic $V(h)$.

The estimation of the bias term relies on iterated kernel estimators. Define

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

The idea is to estimate the bias $\|g - K_h \star g\|^2$ by the supremum of $\|\hat{g}_{h'} - \hat{g}_{h,h'}\|^2$ for h' belonging to an adequate set \mathcal{H} . However, this introduces an additional variance term which must be subtracted and leads to following estimation of the bias term:

$$A(h) = \sup_{h' \in \mathcal{H}} \{ \|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2 - V(h') \}_+, \quad (4.64)$$

with $\mathcal{H} = \{h_j, 1 \leq j \leq M\}$ and M to be specified later. Finally, h is chosen by the following data-driven criterion:

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

Theorem 4.3 *Assume (Ker[1]) to (Ker[3]), (H2-(8))-(H3-g)-(H4-g), and $\int v^2 |g^*(v)|^2 dv := A < +\infty$. Consider \mathcal{H} such that $M \leq n\Delta$, $\forall h \in \mathcal{H}, h \geq 1/(n\Delta)$ and*

$$\forall C > 0, \sum_{h \in \mathcal{H}} h^{-1/2} \exp(-Ch^{-1/2}) \leq \Sigma(C) < +\infty.$$

Then we have

$$\mathbb{E}[\|g - \hat{g}_{\hat{h}}\|^2] \leq C \inf_{h \in \mathcal{H}} \{ \|g - g \star K_h\|^2 + V(h) \} + C' \Delta^2 + C'' \frac{\log^2(n\Delta)}{n\Delta}.$$

Examples of sets \mathcal{H} fitting our assumptions are $\mathcal{H} = \{1/k, k = 1, \dots, [n\Delta]\}$, or $\mathcal{H} = \{2^{-k}, k = 1, \dots, \log([n\Delta])\}$.

Remark 4.7 *The infimum in the bound of Theorem 4.3 means that the estimator $\hat{g}_{\hat{h}}$ automatically reaches the optimal rate stated in Proposition 4.9.*

4.3.4 Proof of Theorem 4.3

The goal is to bound $\mathbb{E}[\|g - \hat{g}_{\hat{h}}\|^2]$. To do this, we fix $h \in \mathcal{H}$ and write

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

The definitions of $A(h)$ and \hat{h} imply:

$$\begin{aligned} \|g - \hat{g}_{\hat{h}}\|^2 &\leq 3\|\hat{g}_{\hat{h}} - \hat{g}_{h,\hat{h}}\|^2 + 3\|\hat{g}_{h,\hat{h}} - \hat{g}_h\|^2 + 3\|\hat{g}_h - g\|^2 \\ &\leq 3(V(\hat{h}) + A(h)) + 3(A(\hat{h}) + V(h)) + 3\|\hat{g}_h - g\|^2 \end{aligned}$$

Again, by definition of \hat{h} , $A(\hat{h}) + V(\hat{h}) \leq A(h) + V(h)$. Therefore, rearranging terms yields

$$\|g - \hat{g}_{\hat{h}}\|^2 \leq 6(A(h) + V(h)) + 3\|\hat{g}_h - g\|^2. \quad (4.65)$$

Consequently,

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

The bound for $\mathbb{E}(\|\hat{g}_h - g\|^2)$ is given by Proposition 4.8. We have to bound $\mathbb{E}[A(h)]$. Let us set $g_{h,h'} = \mathbb{E}[\hat{g}_{h,h'}]$ and $g_h = \mathbb{E}[\hat{g}_h]$. We write,

$$\hat{g}_{h,h'} - \hat{g}_{h'} = \hat{g}_{h,h'} - g_{h,h'} - \hat{g}_{h'} + g_{h'} + g_{h,h'} - g_{h'}, \quad (4.66)$$

and study the last term of the above decomposition:

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

This can be written using notations (4.58)-(4.59)-(4.60),

$$\begin{aligned} |g_{h,h'}(x) - g_{h'}(x)| &= |K_{h'} \star b_h(x) + b_{h',2}(x)| \\ &\leq |K_{h'} \star b_h(x)| + |b_{h',2}(x)| \end{aligned}$$

The Young inequality with $p = 1, r = q = 2$ (see Appendix) and Lemma 4.3 imply:

$$\|g_{h,h'} - g_{h'}\|^2 \leq 2(\|K_{h'} \star b_h\|^2 + \|b_{h',2}\|^2) \leq 2(\|K_{h'}\|_1^2 \|b_h\|^2 + c'_1 \Delta^2), \quad (4.67)$$

where c'_1 is defined in Lemma 4.3 and $\|K_{h'}\|_1 = \|K\|_1$.

Then by inserting (4.67) in decomposition (4.66), we find:

$$\begin{aligned} A(h) &= \sup_{h'} \{ \|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2 - V(h') \}_+ \\ &\leq 3 \sup_{h'} \{ \|\hat{g}_{h,h'} - g_{h,h'}\|^2 - V(h')/6 \}_+ \\ &\quad + 3 \sup_{h'} \{ \|\hat{g}_{h'} - g_{h'}\|^2 - V(h')/6 \}_+ + 6\|K\|_1^2 \|b_h\|^2 + 12c'_1 \Delta^2. \end{aligned} \quad (4.68)$$

The following proposition deals with the first two terms of (4.68).

Proposition 4.10 *Under the assumptions of Theorem 4.3, we have*

$$\mathbb{E} \left[\sup_{h' \in \mathcal{H}} \{ \|\hat{g}_{h'} - g_{h'}\|^2 - V(h')/6 \}_+ \right] \leq \frac{C \log^2(n\Delta)}{n\Delta}, \quad (4.69)$$

and

$$\mathbb{E} \left[\sup_{h' \in \mathcal{H}} \{ \|\hat{g}_{h,h'} - g_{h,h'}\|^2 - V(h')/6 \}_+ \right] \leq \frac{C' \log^2(n\Delta)}{n\Delta} \quad (4.70)$$

Before proving Proposition 4.10, we conclude the proof of Theorem 4.3. Inequalities (4.69) et (4.70) together with (4.68) imply for all $h \in \mathcal{H}$:

$$\mathbb{E}[\|g - \hat{g}_h\|^2] \leq C(\|g - K_h \star g\|^2 + V(h)) + \frac{C' \log^2 n\Delta}{n\Delta} + C'' \Delta^2.$$

So the proof is complete. \square

Proof of Proposition 4.10.

We only prove (4.70) as (4.69) is analogous and slightly simpler. The scheme is similar to the proof of Theorem 4.1. We set $\hat{g}_h = \hat{g}_h^{(1)} + \hat{g}_h^{(2)}$ with

$$\hat{g}_h^{(1)}(x) = \frac{1}{n\Delta} \sum_{j=1}^n Z_j K_h(x - Z_j) \mathbb{I}_{\{|Z_j| \leq k_n \sqrt{\Delta}\}}, \quad (4.71)$$

and $g_h^{(i)} = \mathbb{E}(\hat{g}_h^{(i)})$, $\hat{g}_{h,h'}^{(i)} = K_{h'} \star \hat{g}_h^{(i)}$, $g_{h,h'}^{(i)} = \mathbb{E}(\hat{g}_{h,h'}^{(i)})$ for $i = 1, 2$. Here,

$$k_n = c_0 \frac{\sqrt{n}}{\log(n\Delta)}$$

where c_0 is a constant to be defined. Consequently,

$$\begin{aligned} \mathbb{E} \left[\sup_{h'} \{ \|\hat{g}_{h,h'} - g_{h,h'}\|^2 - V(h')/6 \}_+ \right] &\leq 2\mathbb{E} \left[\sup_{h'} \{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \}_+ \right] \\ &\quad + 2\mathbb{E} \left[\sup_{h'} \|\hat{g}_{h,h'}^{(2)} - g_{h,h'}^{(2)}\|^2 \right] := T_1 + T_2 \end{aligned}$$

We define $\theta_\Delta^{(1)}, \theta_\Delta^{(2)}, \hat{\theta}_\Delta^{(1)}, \hat{\theta}_\Delta^{(2)}$ as in (4.26) and (4.27). Using the relation analogous to (4.54), we have

$$\begin{aligned} \|\hat{g}_{h,h'}^{(2)} - g_{h,h'}^{(2)}\|^2 &= \frac{1}{2\pi\Delta^2} \int |\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)|^2 |K^*(uh)K^*(uh')|^2 du \\ &\leq \frac{\|K\|_1^2}{2\pi\Delta^2} \int |\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)|^2 |K^*(uh)|^2 du \end{aligned}$$

Thus

$$\begin{aligned}
T_2 &= \mathbb{E} \left[\sup_{h'} \|\hat{g}_{h,h'}^{(2)} - g_{h,h'}^{(2)}\|^2 \right] \leq \frac{\|K\|_1^2}{2\pi\Delta^2} \int \mathbb{E} \left[|\hat{\theta}_\Delta^{(2)}(u) - \theta_\Delta^{(2)}(u)|^2 \right] |K^*(uh)|^2 du \\
&\leq \frac{\|K\|_1^2}{2\pi\Delta^2} \int \frac{\mathbb{E}(Z_1^2 \mathbb{I}_{\{|Z_1| > k_n \sqrt{\Delta}\}})}{n} |K^*(uh)|^2 du \\
&\leq \frac{\|K\|_1^2}{2\pi n k_n^2 \Delta^3} \int \mathbb{E}(Z_1^4) |K^*(uh)|^2 du = \frac{\|K\|_1^2 \|K\|^2}{n h k_n^2 \Delta^3} \mathbb{E}(Z_1^4) \\
&\leq \frac{\|K\|_1^2 \|K\|^2}{k_n^2 \Delta} \mathbb{E}(Z_1^4 / \Delta) \leq C \frac{\log^2(n\Delta)}{n\Delta},
\end{aligned}$$

by using the value of k_n . This ends the study of T_2 . Now we consider T_1 and write first

$$\mathbb{E} \left[\sup_{h'} \{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \}_+ \right] \leq \sum_{h' \in \mathcal{H}} \mathbb{E} \left[\{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \}_+ \right].$$

Next we notice

$$\|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 = \sup_{t \in \bar{\mathcal{B}}(1)} \langle \hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}, t \rangle^2$$

where $\bar{\mathcal{B}}(1) = \{t \in \mathbb{L}^2 \cap \mathbb{L}^1(\mathbb{R}), \|t\| = 1\}$. Let $\mathcal{B}(1)$ be a countable subset of $\bar{\mathcal{B}}(1)$ with closure equal to $\bar{\mathcal{B}}(1)$. Then

$$\sup_{t \in \bar{\mathcal{B}}(1)} \langle \hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}, t \rangle^2 = \sup_{t \in \mathcal{B}(1)} \langle \hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}, t \rangle^2$$

and we can apply the Talagrand inequality to the empirical process

$$v_{n,K}(t) = \langle \hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}, t \rangle = \frac{1}{2\pi} \int (\hat{\theta}_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)) K^*(uh) K^*(uh') t^*(u) du.$$

Indeed, $v_{n,K}$ can also be written $v_{n,K}(t) = n^{-1} \sum_{i=1}^n [f_i(Z_i) - \mathbb{E}(f_i(Z_i))]$ with here

$$f_i(z) = \frac{z \mathbb{I}_{\{|z| \leq k_n \sqrt{\Delta}\}}}{2\pi\Delta} \int e^{-ixz} K^*(xh) K^*(xh') t^*(x) dx.$$

(see the proof of Lemma 4.1 where an analogous empirical process is defined). To apply Lemma .1, we compute the three quantities M, H^2 and v . First, for $t \in \mathcal{B}(1)$, we have

$$\sup_{z \in \mathbb{R}} |f_i(z)| \leq \frac{k_n}{2\pi\sqrt{\Delta}} \|t^*\| \|K\|_1 \left(\int |K^*(xh')|^2 dx \right)^{1/2} \leq \|K\|_1 \|K\| \frac{k_n}{\sqrt{h'\Delta}} := M$$

Next, it is clear that

$$\begin{aligned} \mathbb{E} \left(\sup_{t \in \mathcal{B}(1)} [v_{n,K}(t)]^2 \right) &\leq \frac{1}{2\pi\Delta^2} \int \mathbb{E}(|\hat{\theta}_\Delta^{(1)}(u) - \theta_\Delta^{(1)}(u)|^2) |K^*(uh)K^*(uh')|^2 du \\ &\leq \frac{\mathbb{E}(Z_1^2) \|K\|_1^2 \|K\|^2}{nh'\Delta^2} := H^2. \end{aligned}$$

To compute v , we proceed as in the proof of Lemma 4.1. Recall the definitions

$$p_\Delta^*(x) = \mathbb{E}(Z_1^2 1_{\{|Z_1| \leq k_n \sqrt{\Delta}\}} e^{ixZ_1}) = \Delta \int z 1_{\{|z| \leq k_n \sqrt{\Delta}\}} e^{ixz} \mathbb{E}(g(z - Z_1)).$$

We have for all $t \in \mathcal{B}(1)$,

$$\begin{aligned} \text{Var}(f_t(Z_1)) &\leq \frac{1}{4\pi^2\Delta^2} \iint p_\Delta^*(x-y) K^*(-xh) K^*(-xh') t^*(-x) K^*(yh) K^*(yh') t^*(y) dx dy \\ &\leq \frac{\|K\|_1^2}{4\pi^2\Delta^2} \iint |p_\Delta^*(x-y) K^*(-xh') t^*(-x) K^*(yh') t^*(y)| dx dy \\ &\leq \frac{\|K\|_1^2}{4\pi^2\Delta^2} \left(\iint |p_\Delta^*(x-y) K^*(-xh') K^*(yh')|^2 dx dy \iint |t^*(-x) t^*(y)|^2 dx dy \right)^{1/2} \end{aligned}$$

$$\begin{aligned} \text{Var}(f_t(Z_1)) &\leq \frac{\|K\|_1^3}{2\pi\Delta^2} \left(\iint |p_\Delta^*(x-y) K^*(yh')|^2 dx dy \right)^{1/2} \\ &\leq \frac{\|K\|_1^3}{2\pi\Delta^2} \left(\int |p_\Delta^*(z)|^2 dz \int |K^*(yh')|^2 dy \right)^{1/2} \\ &\leq \frac{\|K\|_1^3 \|K\|}{\sqrt{2\pi h'} \Delta^2} \left(\int |p_\Delta^*(z)|^2 dz \right)^{1/2}. \end{aligned}$$

We showed in the proof of Lemma 4.1 that, using Proposition 3.3 and under (H4-g),

$$\int |p_\Delta^*(z)|^2 dz \leq 4\pi\Delta^2 (M_2 + \mathbb{E}(Z_1^2) \|g\|^2).$$

Therefore we get

$$\sup_{t \in \mathcal{B}(1)} \text{Var}(f_t(Z_1)) \leq \frac{\sqrt{2} \|K\|_1^3 \|K\| (M_2 + \mathbb{E}(Z_1^2) \|g\|^2)}{\sqrt{h'} \Delta} := v.$$

Then, setting $V(h')/12 = 4H^2$, we get

$$\begin{aligned} \mathbb{E} \left[\{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \}_+ \right] &= \mathbb{E} \left(\sup_{t \in \mathcal{B}(1)} v_{n,K}^2(t) - 4H^2 \right) \\ &\leq \frac{C_1}{n\Delta} \left(\frac{1}{\sqrt{h'}} e^{-C_2/\sqrt{h'}} + \frac{k_n^2}{nh'} e^{-C_3\sqrt{n}/k_n} \right) \end{aligned}$$

Then if the choice of k_n is such that $c_0 \leq C_3/4$, we obtain

$$\mathbb{E} \left[\left\{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \right\}_+ \right] \leq \frac{C_1}{n\Delta} \left(\frac{1}{\sqrt{h'}} e^{-C_2/\sqrt{h'}} + \frac{1}{(n\Delta)^4 h' \log^2(n\Delta)} \right).$$

Therefore, using the assumptions on \mathcal{H} , i.e. $\text{Card}(\mathcal{H}) \leq n\Delta, \forall h' \in \mathcal{H}, h' \geq 1/(n\Delta)$ and $\sum_{h'} (h')^{-1/2} e^{-C_2/\sqrt{h'}} < +\infty$, if $n\Delta \geq e$, we obtain

$$\mathbb{E} \left[\sup_{h'} \left\{ \|\hat{g}_{h,h'}^{(1)} - g_{h,h'}^{(1)}\|^2 - V(h')/12 \right\}_+ \right] \leq \frac{C}{n\Delta}.$$

The proof of Proposition 4.10 is complete. \square

5 Adaptive estimation with no Gaussian component

In this section, we study the estimation of $\ell(x) = x^2 n(x)$ under (H1- ℓ). We only treat the deconvolution approach and just give below indications on the other two approaches (estimation on a compact set by projection, kernel estimation).

5.1 Deconvolution approach

In addition to (H1- ℓ), we assume:

$$(H3-\ell) \quad \ell \in \mathbb{L}^2(\mathbb{R})$$

$$(H4-\ell) \quad \int x^8 n^2(x) dx = \int x^4 \ell^2(x) dx < \infty.$$

By (H1- ℓ), $\ell \in \mathbb{L}^1(\mathbb{R})$ and the characteristic exponent of the process (L_t) is given by formula (2.7). Assumption (H4- ℓ) is only required for the adaptive result.

5.1.1 Two collections of estimators with cut-off parameter

The deconvolution method requires to define first an estimator of the Fourier transform ℓ^* of ℓ . We propose two estimators $\hat{\ell}^*, \bar{\ell}^*$ of ℓ^* . The former has a smaller bias than the latter but is heavier to implement and more cumbersome to study. Both provide the same variance order. For the first one, we suppose that we have at our disposal a $2n$ -sample, $(Z_k)_{1 \leq k \leq 2n}$, with $Z_k = Z_k^\Delta = L_{k\Delta} - L_{(k-1)\Delta}$. Under (H1- ℓ), φ_Δ is C^2 . Derivating φ_Δ yields

$$\frac{\varphi'_\Delta(u)}{\varphi_\Delta(u)} = \Delta \psi'(u) = i\Delta \left(b + \int \frac{e^{iux} - 1}{x} \ell(x) dx \right) = i\Delta \left(b + i \int_0^u \ell^*(v) dv \right),$$

using $e^{iux} - 1 = ix \int_0^u e^{ivx} dv$. Derivating again yields

$$\ell^*(u) = -\frac{1}{\Delta} \frac{\varphi''_{\Delta}(u)\varphi_{\Delta}(u) - (\varphi'_{\Delta}(u))^2}{\varphi_{\Delta}^2(u)} = -\psi''(u). \quad (5.1)$$

Splitting the $2n$ -sample into two independent subsamples of n observations, we introduce the following empirical unbiased estimators of $\varphi_{\Delta}(u)$, $\varphi'_{\Delta}(u)$, $\varphi''_{\Delta}(u)$:

$$\hat{\varphi}_{\Delta,q}^{(j)}(u) = \frac{1}{n} \sum_{k=1+(q-1)n}^{qn} (iZ_k)^j e^{iuZ_k}, \quad j = 0, 1, 2, \quad q = 1, 2.$$

Considering the expression of ℓ^* in (5.1), we replace φ_{Δ} , φ'_{Δ} , φ''_{Δ} in the numerator by the empirical estimators built on the two independent subsamples of size n . In the denominator, φ_{Δ}^2 is simply replaced by 1. This gives the first estimator of ℓ^* :

$$\hat{\ell}^*(u) = \frac{1}{\Delta} \left(\hat{\varphi}_{\Delta,1}^{(1)}(u)\hat{\varphi}_{\Delta,2}^{(1)}(u) - \hat{\varphi}_{\Delta,1}^{(2)}(u)\hat{\varphi}_{\Delta,2}^{(0)}(u) \right). \quad (5.2)$$

Hence, using independence of the two subsamples,

$$\mathbb{E}\hat{\ell}^*(u) = \ell^*(u) + \ell^*(u)(\varphi_{\Delta}^2(u) - 1). \quad (5.3)$$

Introducing a cut-off parameter m , we define an associated estimator of ℓ

$$\hat{\ell}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \hat{\ell}^*(u) du.$$

This means that $\hat{\ell}_m^*(u) = \hat{\ell}^*(u)1_{[-\pi m, \pi m]}(u)$. By integration, the following expression is available

$$\hat{\ell}_m(x) = \frac{1}{n^2\Delta} \sum_{1 \leq j, k \leq n} (Z_k^2 - Z_k Z_{n+j}) \frac{\sin(\pi m(Z_k + Z_{j+n} - x))}{\pi(Z_k + Z_{j+n} - x)}.$$

This gives a first collection of estimators ($\hat{\ell}_m, m > 0$).

We also define, based on the full sample, the unbiased estimator of φ''_{Δ} :

$$\hat{\varphi}_{\Delta}^{(2)}(u) = \frac{1}{2n} \sum_{k=1}^{2n} (iZ_k)^2 e^{iuZ_k},$$

and set

$$\bar{\ell}^*(u) = -\frac{1}{\Delta} \hat{\varphi}_{\Delta}^{(2)}(u). \quad (5.4)$$

Here, using (5.1), we get

$$\mathbb{E}\bar{\ell}^*(u) = -\frac{1}{\Delta} \varphi''_{\Delta}(u) = \ell^*(u) + \ell^*(u)(\varphi_{\Delta}(u) - 1) - \Delta \varphi_{\Delta}(u)(\psi'(u))^2. \quad (5.5)$$

Thus, $\bar{\ell}^*$ is simpler but has an additional bias term. We set:

$$\bar{\ell}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \bar{\ell}^*(u) du = \frac{1}{2n\Delta} \sum_{k=1}^{2n} Z_k^2 \frac{\sin(\pi m(Z_k - x))}{\pi(Z_k - x)}. \quad (5.6)$$

This gives a second collection of estimators $(\bar{\ell}_m, m > 0)$.

Recall that the characteristic exponent satisfies $\psi'(u) = ib - \int_0^u \ell^*(v) dv$, that we have set $c(u) = |b| + |\int_0^u \ell^*(v) dv|$ and that $|\varphi_\Delta(u) - 1| \leq \Delta |u| c(u)$ (see Proposition 3.4). If ℓ^* is integrable, $c(u) \leq |b| + \|\ell^*\|_1$.

The risk with fixed cut-off parameter is ruled by the following proposition.

Proposition 5.1 *Assume that (H1- ℓ)-(H2-(4)) and (H3- ℓ) hold. Then*

$$\mathbb{E}(\|\hat{\ell}_m - \ell\|^2) \leq \|\ell_m - \ell\|^2 + 72\mathbb{E}(Z_1^4/\Delta) \frac{m}{n\Delta} + \frac{4\Delta^2}{\pi} \int_{-\pi m}^{\pi m} u^2 c^2(u) |\ell^*(u)|^2 du, \quad (5.7)$$

$$\begin{aligned} \mathbb{E}(\|\bar{\ell}_m - \ell\|^2) &\leq \|\ell_m - \ell\|^2 + \mathbb{E}(Z_1^4/\Delta) \frac{m}{n\Delta} \\ &\quad + \frac{2\Delta^2}{\pi} \int_{-\pi m}^{\pi m} u^2 c^2(u) |\ell^*(u)|^2 du + C\Delta^2 B_m, \end{aligned} \quad (5.8)$$

with C a constant, $B_m = (2/\pi) \int_{-\pi m}^{\pi m} |\psi'(u)|^4 du$ satisfies $B_m = O(m)$ if $\ell^* \in \mathbb{L}_1(\mathbb{R})$ and $B_m = O(m^5)$ otherwise.

Proof. The proof follows the same lines as Proposition 4.3 and uses Proposition 3.4. The Parseval formula gives

$$\|\hat{\ell}_m - \ell\|^2 = (1/(2\pi)) \|\hat{\ell}_m^* - \ell^*\|^2.$$

As

$$\ell^*(u) - \ell_m^*(u) = \ell^*(u) \mathbb{I}_{|u| \geq \pi m}$$

is orthogonal to $\hat{\ell}_m^* - \ell_m^*$ which has its support in $[-\pi m, \pi m]$,

$$\|\hat{\ell}_m - \ell\|^2 = \frac{1}{2\pi} (\|\ell^* - \ell_m^*\|^2 + \|\ell_m^* - \hat{\ell}_m^*\|^2).$$

The first term

$$(1/(2\pi)) \|\ell^* - \ell_m^*\|^2 = \|\ell - \ell_m\|^2 = \frac{1}{2\pi} \int_{|u| \geq \pi m} |\ell^*(u)|^2 du$$

is the classical squared bias term. Next,

$$\begin{aligned} \hat{\ell}_m^*(u) - \ell_m^*(u) &= [\hat{\ell}_m^*(u) - \mathbb{E}(\hat{\ell}_m^*(u))] + [\mathbb{E}(\hat{\ell}_m^*(u)) - \ell_m^*(u)] \\ &= [\hat{\ell}_m^*(u) - \mathbb{E}(\hat{\ell}_m^*(u))] + [\varphi_\Delta^2(u) - 1] \ell^*(u) \mathbb{I}_{|u| \leq \pi m}. \end{aligned}$$

Bounding the norm of $\|\hat{\ell}_m^* - \ell_m^*\|^2$ by twice the sum of the norms of the two elements of the decomposition, we get

$$\begin{aligned}\mathbb{E}(\|\hat{\ell}_m - \ell_m\|^2) &\leq \frac{1}{\pi} \mathbb{E} \left(\int_{-\pi m}^{\pi m} |\hat{\ell}^*(u) - \mathbb{E}\hat{\ell}^*(u)|^2 du \right) + \frac{1}{\pi} \int_{-\pi m}^{\pi m} |\varphi_\Delta^2(u) - 1|^2 |\ell^*(u)|^2 du \\ &\leq \frac{1}{\pi} \left(\int_{-\pi m}^{\pi m} \text{Var}(\hat{\ell}^*(u)) du \right) + \frac{4\Delta^2}{\pi} \int_{-\pi m}^{\pi m} u^2 c^2(u) |\ell^*(u)|^2 du\end{aligned}$$

(see Proposition 3.4 for the upper bound of $|\varphi_\Delta(u) - 1|$ and note that $|\varphi_\Delta(u)| \leq 1$). Now, we use the decomposition:

$$\begin{aligned}\Delta(\hat{\ell}^*(u) - \mathbb{E}(\hat{\ell}^*(u))) &= (\hat{\varphi}_{\Delta,1}^{(1)}(u) - \varphi'_\Delta(u))(\hat{\varphi}_{\Delta,2}^{(1)}(u) - \varphi'_\Delta(u)) \\ &\quad + (\hat{\varphi}_{\Delta,1}^{(1)}(u) - \varphi'_\Delta(u))\varphi'_\Delta(u) + (\hat{\varphi}_{\Delta,2}^{(1)}(u) - \varphi'_\Delta(u))\varphi'_\Delta(u) \\ &\quad - (\hat{\varphi}_{\Delta,1}^{(2)}(u) - \varphi''_\Delta(u))(\hat{\varphi}_{\Delta,2}^{(0)}(u) - \varphi_\Delta(u)) \\ &\quad - (\hat{\varphi}_{\Delta,1}^{(2)}(u) - \varphi''_\Delta(u))\varphi_\Delta(u) - (\hat{\varphi}_{\Delta,2}^{(0)}(u) - \varphi_\Delta(u))\varphi''_\Delta(u).\end{aligned}\quad (5.9)$$

Considering each term consecutively and exploiting the independence of the samples, we obtain

$$\begin{aligned}\text{Var}(\hat{\ell}^*(u)) &\leq \frac{6}{\Delta^2} \left(\frac{\mathbb{E}^2(Z_1^2)}{n^2} + 2 \frac{\mathbb{E}^2(Z_1^2)}{n} + \frac{\mathbb{E}(Z_1^4)}{n^2} + 2 \frac{\mathbb{E}(Z_1^4)}{n} \right) \\ &\leq 36 \frac{\mathbb{E}(Z_1^4/\Delta)}{n\Delta}.\end{aligned}\quad (5.10)$$

Thus, (5.7) is proved. Analogously, we have

$$\mathbb{E}(\|\bar{\ell}_m - \ell\|^2) \leq \|\ell_m - \ell\|^2 + \frac{1}{\pi} \int_{-\pi m}^{\pi m} |\mathbb{E}\bar{\ell}^*(u) - \ell^*(u)|^2 du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} \text{Var}(\bar{\ell}^*(u)) du$$

For the variance of $\bar{\ell}^*(u)$, we use: $\bar{\ell}^*(u) - \mathbb{E}\bar{\ell}^*(u) = -\Delta^{-1}(\hat{\varphi}_\Delta^{(2)}(u) - \varphi''_\Delta(u))$. Thus,

$$\text{Var}(\bar{\ell}^*(u)) \leq \frac{1}{2n\Delta} \mathbb{E}(Z_1^4/\Delta).$$

Next, for the bias of $\bar{\ell}^*(u)$, we use (see (5.5)):

$$|\mathbb{E}\bar{\ell}^*(u) - \ell^*(u)|^2 \leq 2|\ell^*(u)|^2 |\varphi_\Delta(u) - 1|^2 + 2\Delta^2 |\psi'(u)|^4.$$

Hence, there is an additional term in the risk bound equal to

$$\frac{2}{\pi} \Delta^2 \int_{-\pi m}^{\pi m} |\psi'(u)|^4 du = \Delta^2 B_m.\quad (5.11)$$

If ℓ^* is integrable, $|\psi'(u)| \leq |b| + \|\ell^*\|_1$, and $B_m = O(m)$. Otherwise, $|\psi'(u)|^4 \leq C|u|^4$, and $B_m = O(m^5)$.

Proposition 5.1 allows to find rates of convergence of the \mathbb{L}^2 -risk of estimators with fixed cut-off parameter m for functions ℓ belonging to Sobolev classes (4.14).

Proposition 5.2 Assume that (H1- ℓ)-(H2-(4)) and (H3- ℓ) hold and that ℓ belongs to $\mathcal{C}(a, L)$ with $a > 1/2$. Consider the asymptotic setting (2.1) and assume that $m \leq n\Delta$. If $n\Delta^2 \leq 1$, then, for the choice $m = O((n\Delta)^{1/(2a+1)})$, we have:

$$\mathbb{E}(\|\hat{\ell}_m - \ell\|^2) \leq O((n\Delta)^{-2a/(2a+1)}).$$

If $a \geq 1$, the condition $n\Delta^2 \leq 1$ can be replaced by $n\Delta^3 \leq 1$.

If $0 < a \leq 1/2$, the constraint on Δ is $n\Delta^{5/3} \leq 1$.

The same results hold for $\bar{\ell}_m$.

Proof. The proof is analogous to the proof of Proposition 4.4. The best compromise between $\|\ell - \ell_m\|^2$ with $\ell \in \mathcal{C}(a, L)$ and $m/(n\Delta)$ leads to $m = O((n\Delta)^{1/(2a+1)})$ and yields the order $O((n\Delta)^{-2a/(2a+1)})$.

It remains to find constraints on Δ implying that the other terms in (5.7)-(5.8) have order less than $O((n\Delta)^{-2a/(2a+1)})$. For $a > 1/2$,

$$|\int_0^u |\ell^*(v)|dv| \leq \sqrt{L \int (1+v^2)^{-a} dv} < +\infty.$$

Therefore, ℓ^* is integrable, $|\psi'(u)| \leq c(u) \leq |b| + \|\ell^*\|_1$ and $B_m = O(m)$.

The last term in the risk bound (5.7) is less than

$$K\Delta^2 \int_{-\pi m}^{\pi m} u^2 |\ell^*(u)|^2 du \leq L\Delta^2 (\pi m)^{2(1-a)_+}.$$

If $a \geq 1$ and $n\Delta^3 \leq 1$, we have $\Delta^2 (\pi m)^{2(1-a)_+} = \Delta^2 \leq (n\Delta)^{-1}$.

If $a \in (1/2, 1)$, the inequality $\Delta^2 m^{2(1-a)} \leq m^{-2a}$ is equivalent to $\Delta^2 m^2 \leq 1$. As $m \leq n\Delta$, $\Delta^2 m^2 \leq 1$ holds if $n\Delta^2 \leq 1$.

For the additional bias term appearing in the risk bound of $\bar{\ell}_m$, we are in the case $B_m = O(m)$. Thus, for $m = O((n\Delta)^{1/(2a+1)})$, $m\Delta^2 \leq m^{-2a}$ holds if $m^{1+2a}\Delta^2 = (n\Delta)\Delta^2 \leq 1$ which in turn holds if $n\Delta^3 \leq 1$.

If $a \leq 1/2$,

$$|\int_0^u |\ell^*(v)|dv| = O(|u|^{1/2-a}).$$

Hence, the last term in (5.7) is of order $\Delta^2 m^{3-4a}$ which is less than m^{-2a} if $\Delta^2 m^{3-2a} \leq 1$ and thus $\Delta^2 m^3 \leq 1$. This requires $n\Delta^{5/3} \leq 1$. The same holds for $\bar{\ell}_m$.

5.1.2 Data-driven choice of the bandwidth and adaptive estimator

We describe now how to choose m in a data-driven way leading to an adaptive estimator, *i.e.* attaining automatically the optimal rate of convergence without knowledge of the regularity of the unknown function ℓ . Recall the collection of subspaces (S_m) of $\mathbb{L}^2(\mathbb{R})$ defined in (4.15) where each space S_m is generated by the orthonormal basis (4.19).

For a function $t \in \mathcal{S}_m$, define

$$\Gamma_n^{(1)}(t) = \|t\|^2 - \frac{1}{\pi} \langle \hat{\ell}^*, t^* \rangle = \|t\|^2 - 2 \langle \hat{\ell}_m, t \rangle,$$

so that

$$\hat{\ell}_m = \arg \min_{t \in \mathcal{S}_m} \Gamma_n^{(1)}(t),$$

and $\Gamma_n^{(1)}(\hat{\ell}_m) = -\|\hat{\ell}_m\|^2$. In the same way, we set

$$\Gamma_n^{(2)}(t) = \|t\|^2 - \frac{1}{\pi} \langle \bar{\ell}^*, t^* \rangle = \|t\|^2 - 2 \langle \bar{\ell}_m, t \rangle,$$

and

$$\bar{\ell}_m = \arg \min_{t \in \mathcal{S}_m} \Gamma_n^{(2)}(t),$$

Explicit expressions of $\|\hat{\ell}_m\|^2$ and $\|\bar{\ell}_m\|^2$ are available. We give the formula for $\|\bar{\ell}_m\|^2$ which is less cumbersome than $\|\hat{\ell}_m\|^2$:

$$\|\bar{\ell}_m\|^2 = \frac{m}{4n^2\Delta^2} \sum_{1 \leq k, l \leq 2n} Z_k^2 Z_l^2 \phi(m(Z_k - Z_l)). \quad (5.12)$$

Now, we need to select m in $\mathcal{M}_n = \{m \in \mathbb{N}, 1 \leq m \leq n\Delta\} = \{1, \dots, m_n\}$. For the estimators $\hat{\ell}_m$, we define

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} (-\|\hat{\ell}_m\|^2 + \text{pen}(m)) \quad (5.13)$$

with

$$\text{pen}(m) = \kappa \frac{m}{n\Delta^2} \left(\left(\frac{1}{n} \sum_{k=1}^n Z_k^2 \right) \left(\frac{1}{n} \sum_{k=n+1}^{2n} Z_k^2 \right) + \frac{1}{n} \sum_{k=1}^n Z_k^4 \right).$$

For the estimators $\bar{\ell}_m$, we define

$$\bar{m} = \arg \min_{m \in \mathcal{M}_n} \left(-\|\bar{\ell}_m\|^2 + \kappa' \frac{m}{n\Delta^2} \left(\frac{1}{2n} \sum_{k=1}^{2n} Z_k^4 \right) \right). \quad (5.14)$$

The following result shows that the above data-driven choices of the cut-off parameter lead to an automatic optimization of the risk.

Theorem 5.1 *Assume (H1- ℓ)-(H2-(16))-(H3- ℓ)-(H4- ℓ). If, moreover, $\ell^* \in \mathbb{L}^1(\mathbb{R})$ and $n\Delta^3 \leq 1$, there exist numerical constants κ, κ' such that*

$$\begin{aligned} \mathbb{E}(\|\hat{\ell}_{\hat{m}} - \ell\|^2) &\leq C \inf_{m \in \mathcal{M}_n} \left(\|\ell - \ell_m\|^2 + \kappa(\Delta \mathbb{E}^2\left(\frac{Z_1^2}{\Delta}\right) + \mathbb{E}\left(\frac{Z_1^4}{\Delta}\right)) \frac{m}{n\Delta} \right) \\ &\quad + \frac{\Delta^2}{\pi} \int_{-\pi m_n}^{\pi m_n} u^2 |\ell^*(u)|^2 du + C \frac{\log^2(n\Delta)}{n\Delta}, \end{aligned}$$

and

$$\mathbb{E}(\|\bar{\ell}_m - \ell\|^2) \leq C \inf_{m \in \mathcal{M}_n} \left(\|\ell - \ell_m\|^2 + \kappa' \mathbb{E} \left(\frac{Z_1^4}{\Delta} \right) \frac{m}{n\Delta} \right) + \frac{\Delta^2}{\pi} \int_{-\pi m_n}^{\pi m_n} u^2 |\ell^*(u)|^2 du + \Delta^2 B_{m_n} + C \frac{\log^2(n\Delta)}{n\Delta},$$

where $B_{m_n} = O(m_n)$ (B_{m_n} is defined in Proposition 5.1).

The proof of Theorem 5.1 follows the same steps as Theorem 4.1 (with some more technical developments) and is therefore omitted. We refer to [20] (Theorem 3.1) for details. By computations analogous to those in the proof of Proposition 5.2, we obtain the following Corollary.

Corollary 5.1 *Assume that the assumptions of Theorem 5.1 are fulfilled. If, for some positive L , $\ell \in \mathcal{C}(a, L)$ with $a > 1/2$, then $\mathbb{E}(\|\hat{\ell}_m - \ell\|^2) = O((n\Delta)^{-2a/(2a+1)})$ provided that $n\Delta^2 \leq 1$. The same holds for $\mathbb{E}(\|\bar{\ell}_m - \ell\|^2)$. If $a \geq 1$, the constraint $n\Delta^3 \leq 1$ is enough.*

5.2 Projection and kernel

Consider a set of n observations (Z_k) . It is possible to use the fact that

$$\hat{\mu}_n^{(2)} = \frac{1}{n\Delta} \sum_{k=1}^n Z_k^2 \delta_{Z_k}$$

approximates the measure $\mu^{(2)}(dx) = \ell(x)dx$. This allows to build as in Section 4.2 and Section 4.3 either estimators of $\ell(\cdot)$ on a compact set A or kernel estimators of $\ell(\cdot)$.

6 Adaptive estimation in the general case

Finally, we study the estimation of $p(x) = x^3 n(x)$ under (H1-p) and in addition

(H3-p) $p \in \mathbb{L}^2(\mathbb{R})$

(H4-p) $\int x^{12} n^2(x) dx = \int x^6 p^2(x) dx < \infty$.

We construct estimators analogous to $\bar{\ell}_m$ based on a sample of size n , $(Z_k)_{1 \leq k \leq n}$, $Z_k = L_{k\Delta} - L_{(k-1)\Delta}$. For this, we need to compute the third derivative of the characteristic function $\varphi_\Delta(u) = \exp \Delta \psi(u)$ where the characteristic exponent $\psi(u)$ is given by formula (2.9):

$$\varphi_\Delta^{(3)}(u) = \varphi_\Delta(u) [\Delta \psi^{(3)}(u) + 3\Delta^2 \psi'(u) \psi''(u) + \Delta^3 (\psi'(u))^3] \quad (6.1)$$

with

$$\begin{aligned}\psi'(u) &= ib - u\sigma^2 + \int ix(e^{iux} - 1)n(x)dx = ib - u\sigma^2 - \int_0^u \ell^*(v)dv, \\ \psi''(u) &= -\sigma^2 - \ell^*(u), \\ \psi^{(3)}(u) &= -ip^*(u).\end{aligned}$$

It follows that:

$$\frac{i}{\Delta}\varphi_{\Delta}^{(3)}(u) = p^*(u) + p^*(u)(\varphi_{\Delta}(u) - 1) + i\varphi_{\Delta}(u)[3\Delta\psi'(u)\psi''(u) + \Delta^2(\psi'(u))^3]$$

The Fourier transform p^* of p is simply estimated by

$$\bar{p}^*(u) = \frac{i}{\Delta}\hat{\varphi}_{\Delta}^{(3)}(u) \quad \text{with} \quad \hat{\varphi}_{\Delta}^{(3)}(u) = \frac{1}{n}\sum_{k=1}^n (iZ_k)^3 e^{iuZ_k}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\bar{p}^*(u) - p^*(u) &= (\varphi_{\Delta}(u) - 1)p^*(u) + 3i\Delta\varphi_{\Delta}(u)\psi'(u)\psi''(u) \\ &\quad + i\Delta^2\varphi_{\Delta}(u)(\psi'(u))^3.\end{aligned}\tag{6.2}$$

By Fourier inversion, we obtain a collection of estimators with cut-off parameter m :

$$\bar{p}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \bar{p}^*(u) du = \frac{1}{n\Delta} \sum_{k=1}^n Z_k^3 \frac{\sin(\pi m(Z_k - x))}{\pi(Z_k - x)}.\tag{6.3}$$

The risk is bounded as follows.

Proposition 6.1 *Under (H1-p)-(H2)(6) and (H3-p),*

$$\begin{aligned}\mathbb{E}(\|\bar{p}_m - p\|^2) &\leq \|p - p_m\|^2 + \mathbb{E}(Z_1^6/\Delta) \frac{m}{n\Delta} \\ &\quad + C(\Delta^2 \int_{-\pi m}^{\pi m} u^2(1+u^2)|p^*(u)|^2 du + \Delta^2 m^3 + \Delta^4 m^7),\end{aligned}\tag{6.4}$$

where $p_m(x) = (2\pi)^{-1} \int_{-\pi m}^{\pi m} e^{-iux} p^*(u) du$.

Proof. As previously, $\|\bar{p}_m - p\|^2 = \frac{1}{2\pi}(\|p^* - p_m^*\|^2 + \|p_m^* - \bar{p}_m^*\|^2)$. The variance of \bar{p}_m satisfies

$$\mathbb{E}(\|\bar{p}_m - p_m\|^2) = \frac{1}{2\pi} \mathbb{E}(\|\bar{p}_m^* - p_m^*\|^2) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} (\text{Var}(\bar{p}^*(u)) + |\mathbb{E}(\bar{p}^*(u)) - p^*(u)|^2) du,$$

where

$$\text{Var}(\bar{p}^*(u)) \leq \frac{\mathbb{E}(Z_1^6)}{n\Delta^2} = \frac{\mathbb{E}(Z_1^6/\Delta)}{n\Delta}.$$

We have $|\ell^*(u)| \leq \|\ell\|_1 < +\infty$. Thus, $|\psi'(u)| \leq |b| + \sigma^2 + \|\ell\|_1$, $|\psi''(u)| \leq \sigma^2 + \|\ell\|_1$ and by Proposition 3.4, $|\varphi_{\Delta}(u) - 1| \leq C\Delta|u|(1+|u|)$. Inserting these bounds in (6.2)

implies

$$|\mathbb{E}(\bar{p}^*(u)) - p^*(u)| \leq C\Delta |p^*(u)| |u| (1 + |u|) + C'\Delta (1 + |u|) + C''\Delta^2 (1 + |u|)^3 \quad (6.5)$$

Gathering the terms gives the announced bound for the risk of \bar{p}_m .

We can state the result analogous to the one of Proposition 5.2.

Proposition 6.2 *Assume that (H1-p), (H2-(6)), (H3-p) hold and that p belongs to $\mathcal{C}(a, L)$. If $n\Delta^{11/7} \leq 1$, then*

$$\mathbb{E}(\|\bar{p}_m - p\|^2) \leq O((n\Delta)^{-2a/(2a+1)}).$$

If $a \geq 1/2$, the condition $n\Delta^{7/5} \leq 1$ can be replaced by $n\Delta^2 \leq 1$.

For the data driven selection of m , we must impose here a restricted collection of models:

$$M_n = \{m \in \mathbb{N}/\{0\}, m \leq \sqrt{n\Delta} := \mu_n\},$$

and set

$$\bar{m} = \arg \min_{m \in M_n} (-\|\bar{p}_m\|^2 + \overline{\text{pen}}(m)) \quad \text{with} \quad \overline{\text{pen}}(m) = \kappa \frac{m}{n\Delta^2} \left(\frac{1}{n} \sum_{k=1}^n Z_k^6 \right). \quad (6.6)$$

The estimator $\bar{p}_{\bar{m}}$ satisfies:

Theorem 6.1 *Assume (H1-p), (H2-(24)), (H3-p), (H4-p) and $n\Delta^2 \leq 1$. Then, there exists a numerical constant κ such that (with $\mu_n = \sqrt{n\Delta}$)*

$$\begin{aligned} \mathbb{E}(\|\bar{p}_{\bar{m}} - p\|^2) &\leq C \inf_{m \in M_n} \left(\|p - p_m\|^2 + \kappa \mathbb{E} \left(\frac{Z_1^6}{\Delta} \right) \frac{m}{n\Delta} \right) \\ &+ C \left(\frac{\Delta^2}{\pi} \int_{-\pi\mu_n}^{\pi\mu_n} u^2 (1 + u^2) |p^*(u)|^2 du + \Delta^2 \mu_n^3 + \Delta^4 \mu_n^7 + \frac{\log^2(n\Delta)}{n\Delta} \right). \end{aligned}$$

For the proof, we refer to [20] (Theorem 4.1).

Remark 6.1 *We could also build other kinds of estimators using the fact that*

$$\hat{\mu}_n^{(3)} = \frac{1}{n\Delta} \sum_{k=1}^n Z_k^3 \delta_{Z_k}$$

approximates the measure $\mu^{(3)}(dx) = p(x)dx$.

7 Drift and Gaussian component estimation

Consider the general case where the observed process is $L_t = bt + \sigma W_t + X_t$ with (X_t) a centered square integrable pure-jump martingale: $X_t = \int_{]0,t]} \int_{\mathbb{R}/\{0\}} x(\hat{p}(du, dx) -$

$du n(x)dx$, and $\hat{p}(du, dx)$ is the random Poisson measure associated with the jumps of (L_t) (or (X_t)) (see (2.10)). By using empirical means of the data Z_k^l (recall that $Z_k = L_{k\Delta} - L_{(k-1)\Delta}$) it is possible to obtain consistent and asymptotically Gaussian estimators of b ($l = 1$) and, under suitable integrability assumptions on the Lévy density, of $\int x^l n(x)dx$ for $l \geq 3$. But this method fails to estimate σ for $l = 2$.

7.1 Empirical means

Consider a Lévy process (L_t) and set $Z_k = L_{k\Delta} - L_{(k-1)\Delta}$ as above. Let us define the empirical means:

$$\hat{b} = \frac{1}{n\Delta} \sum_{k=1}^n Z_k, \quad \hat{c}_l = \frac{1}{n\Delta} \sum_{k=1}^n Z_k^l \text{ for } l \geq 2. \quad (7.1)$$

We prove now that $\hat{b}, \hat{c}_l, l \geq 2$ are consistent and asymptotically Gaussian estimators of the quantities $b, c_l, l \geq 2$ where

$$c_2 = \sigma^2 + \int x^2 n(x)dx, \quad c_l = \int x^l n(x)dx, \quad \text{for } l \geq 3.$$

Proposition 7.1 *Assume that $\int x^2 n(x)dx < +\infty$ and the asymptotic framework (2.1).*

(i) *Under (H2-(2 + ε)) for some positive ε ,*

$$\sqrt{n\Delta}(\hat{b} - b) \text{ converges in distribution to } \mathcal{N}(0, c_2).$$

(ii) *Under (H2-(2(l + ε))) for some positive ε , and if $n\Delta^3$ tends to 0, $\sqrt{n\Delta}(\hat{c}_l - c_l)$ converges in distribution to $\mathcal{N}(0, c_{2l})$.*

Proof. By Proposition 3.1, $\mathbb{E}(Z_k) = \Delta b$ and, for $l \geq 2$, $\mathbb{E}(Z_k^l) = \Delta c_l + o(\Delta)$. Therefore, \hat{b} is an unbiased estimator of b . For $l \geq 2$, $\sqrt{n\Delta}|\mathbb{E}\hat{c}_l - c_l| = \sqrt{n\Delta}O(\Delta)$ which tends to 0 under the additional condition $n\Delta^3 = o(1)$.

Setting $c_1 = b$, $\hat{c}_1 = \hat{b}$, as $\text{Var}Z_k^l = \Delta c_{2l} + o(\Delta)$ for $l \geq 1$, we have $n\Delta \text{Var}\hat{c}_l = c_{2l} + O(\Delta)$. Writing

$$\sqrt{n\Delta}(\hat{c}_l - \mathbb{E}\hat{c}_l) = (n\Delta)^{-1/2} \sum_{k=1}^n (Z_k^l - \mathbb{E}Z_k^l) = \sum_{k=1}^n \chi_{k,n},$$

it is now enough to prove that $\sum_{k=1}^n \mathbb{E}|\chi_{k,n}|^{2+\varepsilon}$ tends to 0. By the moment assumption (H2-(2(l + ε))), we have

$$\sum_{k=1}^n \mathbb{E}|\chi_{k,n}|^{2+\varepsilon} \leq \frac{C}{n^{\varepsilon/2} \Delta^{1+\varepsilon/2}} \left(\mathbb{E}|Z_k|^{l(2+\varepsilon)} + |\mathbb{E}(Z_k^l)|^{2+\varepsilon} \right) \leq \frac{C}{(n\Delta)^{\varepsilon/2}},$$

which gives the result.

7.2 Estimation of the Gaussian component parameter with power variations

Estimators of σ based on power variations of (L_t) have been proposed and mostly studied in the case where $n\Delta = 1$, see [5], [63], [40]. They are studied under the asymptotic framework (2.1) in [1] and [20]. Consider the family of estimators of σ given by

$$\hat{\sigma}(r) = [\hat{\sigma}_n^{(r)}]^{1/r} \quad \text{with} \quad \hat{\sigma}_n^{(r)} = \frac{1}{m_r n \Delta^{r/2}} \sum_{k=1}^n |Z_k|^r, \quad (7.2)$$

where $m_r = \mathbb{E}|X|^r$ for X a standard Gaussian variable. The following result concerns only restricted cases.

Proposition 7.2 *Consider the asymptotic framework (2.1) and assume that $r < 1$ and $n\Delta^{2-r} = o(1)$. Then, $\sqrt{n}(\hat{\sigma}_n^{(r)} - \sigma^r)$ converges in distribution to a $\mathcal{N}(0, \sigma^{2r}(m_{2r}/m_r^2 - 1))$ for:*

(i) $(L_t = bt + \sigma W_t + \Gamma_t)$ where Γ_t is a pure jump process satisfying (H1-g) and

$$\int_{|x| \leq 1} |x|^r n(x) dx < \infty$$

(ii) $(L_t = bt + \sigma W_t + X_t)$, with $X_t = B_{\Gamma_t}$, where W, B, Γ are independent processes, W, B are Brownian motions, Γ is a subordinator with Lévy measure n_Γ satisfying

$$\int_0^{+\infty} \gamma^{r/2} n_\Gamma(\gamma) d\gamma < \infty.$$

In each case, $\sqrt{n}(\hat{\sigma}(r) - \sigma)$ converges in distribution to a $\mathcal{N}(0, (\sigma^2/r^2)(m_{2r}/m_r^2 - 1))$.

Remark 7.1 *It is worth noting that the rate of convergence is \sqrt{n} and not $\sqrt{n\Delta}$. For $r = 1$, the estimator $\hat{\sigma}_n^{(1)}$ is consistent but not asymptotically Gaussian.*

Proof. The study of (7.2) relies on the following result which is standard for $r = 2$.

Lemma 7.1 *Let $Y_t = \theta t + \sigma W_t$ for θ a constant and consider*

$$\tilde{\sigma}_n^{(r)} = \frac{1}{m_r n \Delta^{r/2}} \sum_{k=1}^n |Y_{k\Delta} - Y_{(k-1)\Delta}|^r.$$

Then, for all r , $\sqrt{n}(\tilde{\sigma}_n^{(r)} - \sigma^r)$ converges in distribution to a centered Gaussian distribution with variance $\sigma^{2r}(m_{2r}/m_r^2 - 1)$ as n tends to infinity, Δ tends to 0, $n\Delta$ tends to infinity, and $n\Delta^2$ tends to 0.

Proof of (i). Using that, for $r \leq 1$, $|\sum a_i + b_i|^r - |\sum a_i|^r \leq \sum |b_i|^r$, we get

$$|\hat{\sigma}_n^{(r)} - \tilde{\sigma}_n^{(r)}| \leq \frac{1}{m_r n \Delta^{r/2}} \sum_{k=1}^n |\Gamma_{k\Delta} - \Gamma_{(k-1)\Delta}|^r,$$

where $\tilde{\sigma}_n^{(r)}$ is built with $Y_t = bt + \sigma W_t$ as in Lemma 7.1. Thus, applying Proposition 3.2 (2),

$$\mathbb{E}\sqrt{n}|\hat{\sigma}_n^{(r)} - \tilde{\sigma}_n^{(r)}| \leq \frac{1}{m_r} \sqrt{n} \Delta^{1-r/2} \int |x|^r n(x) dx.$$

Since $r < 1$, the constraint $n\Delta^{2-r} = o(1)$ can be fulfilled and implies $n\Delta^2 = o(1)$. Hence, the result follows from Lemma 7.1.

Proof of (ii). The proof is analogous to the previous one (using Proposition 3.2 (3)) and is omitted.

As $\hat{\sigma}(r) = [\hat{\sigma}_n^{(r)}]^{1/r}$, we conclude for $\hat{\sigma}(r)$ by using the delta-method.

Proof of Lemma 7.1. We have $\mathbb{E}\tilde{\sigma}_n^{(r)} = \frac{1}{m_r} \mathbb{E}|\theta\sqrt{\Delta} + \sigma X|^r$, for X a standard Gaussian variable. Thus

$$\mathbb{E}\tilde{\sigma}_n^{(r)} - \sigma^r = \sigma^r \left(e^{-\theta^2\Delta/2\sigma^2} - 1 \right) + \frac{1}{m_r} e^{-\theta^2\Delta/2\sigma^2} \int |u|^r (e^{\theta u\sqrt{\Delta}/\sigma^2} - 1) e^{-\frac{u^2}{2\sigma^2}} \frac{du}{\sigma\sqrt{2\pi}}.$$

Noting that $e^{\theta u\sqrt{\Delta}/\sigma^2} - 1 = \theta u\sqrt{\Delta}/\sigma^2 + \Delta \sum_{n \geq 2} \frac{1}{n!} (\theta u/\sigma^2)^n \Delta^{n/2-1}$ and that $\int |u|^r u e^{-\frac{u^2}{2\sigma^2}} du / (\sigma\sqrt{2\pi}) = 0$, we obtain

$$|\mathbb{E}\tilde{\sigma}_n^{(r)} - \sigma^r| \leq c\Delta$$

Thus, $\sqrt{n}|\mathbb{E}\tilde{\sigma}_n^{(r)} - \sigma^r| = o(1)$ if $\sqrt{n}\Delta = (n\Delta^2)^{1/2} = o(1)$. Noting that $\mathbb{E}|\theta\sqrt{\Delta} + \sigma X|^k$ converges to $\sigma^k m_k$ as Δ tends to 0, we get $n\text{Var}\tilde{\sigma}_n^{(r)} \rightarrow \sigma^{2r}(m_{2r}/m_r^2 - 1)$.

Finally, we look at $\chi_{k,n} = n^{-1} \left(|\theta\sqrt{\Delta} + \sigma(W_{k\Delta} - W_{(k-1)\Delta})|/\sqrt{\Delta} \right)^r - \mathbb{E}|\theta\sqrt{\Delta} + \sigma X|^r$,

which satisfies $n\mathbb{E}\chi_{k,n}^4 \leq c/n^3$. Hence, $\sqrt{n}(\tilde{\sigma}_n^{(r)} - \mathbb{E}\tilde{\sigma}_n^{(r)})$ converges in distribution to the centered Gaussian with the announced variance which completes the proof. \square

8 Rates of convergence on examples

In this section, we illustrate on examples the possible rates of convergence of the estimators of g and p obtained by Proposition 4.3, Theorem 4.1, Proposition 4.6 and Theorem 4.2 for the estimation of g , Proposition 6.1, Theorem 6.1 for the estimation of p .

8.1 Pure-jump case

The discussion on rates of convergence is different according to the estimation method.

8.1.1 Rates for the Fourier method on examples

We consider models for which (H1-g) holds.

Example 1. Compound Poisson processes.

Let $L_t = \sum_{i=1}^{N_t} Y_i$, where (N_t) is a Poisson process with constant intensity c 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 compound Poisson process with characteristic function (2.5) with $n(x) = cf(x)$ (integrable). Assumptions (H1-g)-(H2-(l)) are equivalent to $e(|Y_1|^l) < \infty$. Assumption (H3-g) is equivalent to $\int_{\mathbb{R}} x^2 f^2(x) dx < \infty$, which holds for instance if $\sup_x f(x) < +\infty$ and $\mathbb{E}(Y_1^2) < +\infty$. The distribution of $Z_1 = L_\Delta$ is:

$$P_\Delta(dz) = P_{Z_1}(dz) = e^{-c\Delta} \left(\delta_0(dz) + \sum_{n \geq 1} f^{*n}(z) \frac{(c\Delta)^n}{n!} dz \right). \quad (8.1)$$

Hence,

$$\mu_\Delta^{(1)}(dz) = e^{-c\Delta} \left(czf(z)dz + c^2\Delta z \sum_{n \geq 2} \frac{c^{n-2}\Delta^{n-2}}{n!} f^{*n}(z)dz \right) \quad (8.2)$$

As f is any density and $g(x) = cx f(x)$, any type of rate can be obtained. Table 1 summarizes the rates obtained when f is Gaussian, exponential or uniform.

Density f	Gaussian $\mathcal{N}(0, 1)$	Exponential $\mathcal{E}(1)$	Uniform $\mathcal{U}([0, 1])$
$g(x) (= cx f(x)) =$	$cxe^{-x^2/2}/\sqrt{2\pi}$	$cxe^{-x} \mathbb{1}_{\mathbb{R}^+}(x)$	$cx \mathbb{1}_{[0,1]}(x)$
$g^*(u) =$	$ciue^{-u^2/2}$	$c/(1-iu)^2$	$c \frac{e^{iu} - 1 - iue^{iu}}{u^2}$
$\int_{ u \geq \pi m} g^*(u) ^2 du =$	$O(me^{-\pi^2 m^2})$	$O(m^{-3})$	$O(m^{-1})$
$\int_{ u \leq \pi m_n} u^2 g^*(u) ^2 du =$	$O(1)$	$O(1)$	$O(m_n)$
Constraint on Δ	$n\Delta^3 \leq 1$	$n\Delta^3 \leq 1$	$n\Delta^2 \leq 1$
Selected $m =$	$m = \sqrt{\log(n\Delta)}/\pi$	$m = O((n\Delta)^{1/4})$	$m = O((n\Delta)^{1/2})$
Rate =	$O\left(\frac{\sqrt{\log(n\Delta)}}{n\Delta}\right)$	$O((n\Delta)^{-3/4})$	$O((n\Delta)^{-1/2})$

Table 1 Choice of m and rates in three compound Poisson examples ($m_n \leq n\Delta$).

For instance, for $\Delta = n^{-a}$, with $a \in [1/3, 1[$, the best risk is of order $\log^{1/2}(n)/n^{2/3}$ in the Gaussian case and of order $n^{-1/2}$ in the exponential case. In the uniform case for $\Delta = n^{-a}$ and now $a \in [1/2, 1[$, the best risk is of order $n^{-1/4}$.

Example 2. The Lévy Gamma process. Let $\alpha > 0, \beta > 0$. The Lévy Gamma process (L_t) with parameters (β, α) is a subordinator (increasing Lévy process) such that, for all $t > 0$, L_t has distribution Gamma with parameters $(\beta t, \alpha)$, *i.e.* has density:

$$\frac{\alpha^{\beta t}}{\Gamma(\beta t)} x^{\beta t - 1} e^{-\alpha x} \mathbf{1}_{x \geq 0}. \quad (8.3)$$

The characteristic function of Z_1 is equal to:

$$\varphi_{\Delta}(u) = \left(\frac{\alpha}{\alpha - iu} \right)^{\beta \Delta}. \quad (8.4)$$

The Lévy density is $n(x) = \beta x^{-1} e^{-\alpha x} \mathbf{1}_{\{x > 0\}}$ so that $g(x) = \beta e^{-\alpha x} \mathbf{1}_{\{x > 0\}}$ satisfies our assumptions. We have: $g^*(u) = \beta / (\alpha - iu)$. Table 2 gives the rate of the risk bound and auxiliary quantities.

Example 2 (continued): Lévy δ process. To illustrate other possibilities of rates, consider a pure jump Lévy process (L_t) with parameters (δ, β, c) and Lévy density

$$n(x) = cx^{\delta - 1/2} x^{-1} e^{-\beta x} \mathbf{1}_{x > 0}.$$

Assumption (H1-g) holds for $\delta > -1/2$. For $\delta > 1/2$, $\int_0^{+\infty} n(x) dx < +\infty$, the process is a compound Poisson process.

For $0 < \delta \leq 1/2$, $\int_0^{+\infty} n(x) dx = +\infty$ and $g(x) = xn(x)$ belongs to $\mathbb{L}^2(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$. This includes the case $\delta = 1/2$ of the Lévy Gamma process.

The case $-1/2 < \delta \leq 0$ and in particular $\delta = 0$, which corresponds to the inverse Gaussian Lévy process, does not fit in this part.

We have:

$$g^*(u) = c \frac{\Gamma(\delta + 1/2)}{(\beta - iu)^{\delta + 1/2}}.$$

Table 2 shows that for $\Delta = n^{-a}$, with $a \in [1/2, 1[$, the best risk is of order $n^{-\delta/(2\delta+1)}$.

Example 3. The variance Gamma stochastic volatility model (see [52]).

Let (W_t) be a Brownian motion, and let (V_t) be a Lévy Gamma process, independent of (W_t) . Assume that the observed process is $L_t = W_{V_t}$. The characteristic function is given by:

$$\varphi_{\Delta}(u) = \mathbb{E}(e^{iuL_{\Delta}}) = \mathbb{E}(e^{-\frac{u^2}{2}V_{\Delta}}) = \left(\frac{\alpha}{\alpha + \frac{u^2}{2}} \right)^{\Delta \beta}.$$

The Lévy measure of (L_t) is equal to:

$$n_L(x) = \beta(2\alpha)^{1/4} |x|^{-1} \exp(-(2\alpha)^{1/2} |x|).$$

The density of $L_{\Delta} = Z_1$ can be computed as it is a variance mixture of Gaussian distributions with mixing distribution Gamma $\Gamma(\beta\Delta, \alpha)$:

$$\begin{aligned} f_{Z_1}(x) &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} v^{\beta\Delta-3/2} e^{-\frac{1}{2}(x^2/v+2\alpha v)} \frac{\alpha^{\beta\Delta}}{\Gamma(\beta\Delta)} dv \\ &= \frac{2}{\sqrt{2\pi}} \frac{\alpha^{\beta\Delta}}{\Gamma(\beta\Delta)} \left(\frac{(2\alpha)^{1/2}}{|x|}\right)^{\frac{1}{2}-\beta\Delta} K_{\beta\Delta-\frac{1}{2}}((2\alpha)^{1/2}|x|) \end{aligned}$$

where K_ν is the modified Bessel function (third kind) with index ν (see *e.g.* [51]).

Now with $\tilde{\alpha} = (2\alpha)^{1/2}$, $\tilde{\beta} = \beta(2\alpha)^{1/4}$,

$$g(x) = \tilde{\beta} \exp(-\tilde{\alpha}x) \mathbb{I}_{x \geq 0} - \tilde{\beta} \exp(\tilde{\alpha}x) \mathbb{I}_{x < 0}, \quad g^*(u) = \frac{2i\tilde{\alpha}\tilde{\beta}u}{\tilde{\alpha}^2 + u^2}.$$

Example 3 (continued). The variance Gamma stochastic volatility model is a special case of bilateral Gamma process (see [50]). Consider the Lévy process L_t with characteristic function

$$\varphi_t(u) = \left(\frac{\alpha}{\alpha - iu}\right)^{\beta t} \left(\frac{\alpha'}{\alpha' + iu}\right)^{\beta' t}$$

and Lévy density

$$n(x) = |x|^{-1} (\beta e^{-\alpha x} \mathbb{I}_{(0,+\infty)}(x) + \beta' e^{-\alpha|x|} \mathbb{I}_{(-\infty,0)}(x)).$$

Rates are given in Table (2).

Process	Example 2	Ex.2 (continued) $\delta \in]0, 1/2[$	Example 3 (continued)
$g^*(u) =$	$\frac{\beta}{\alpha - iu}$	$c \frac{\Gamma(\delta + 1/2)}{(\beta - iu)^{\delta+1/2}}$	$\frac{\beta}{\alpha - iu} - \frac{\beta'}{\alpha' - iu}$
$\int_{ u \geq \pi m} g^*(u) ^2 du =$	$O(1/m)$	$O(1/m^{2\delta})$	$O(1/m)$
$\int_{ u \leq \pi m_n} u^2 g^*(u) ^2 du =$	$O(m_n)$	$O(m_n^{2-2\delta})$	$O(m_n)$
Constraint on Δ	$n\Delta^2 \leq 1$	$n\Delta^2 \leq 1$	$n\Delta^2 \leq 1$
Selected $m =$	$O((n\Delta)^{1/2})$	$O((n\Delta)^{1/(2\delta+1)})$	$O((n\Delta)^{1/2})$
Rate	$O((n\Delta)^{-1/2})$	$O((n\Delta)^{-2\delta/(2\delta+1)})$	$O((n\Delta)^{-1/2})$

Table 2 Choice of m and rates in examples 2, 2 (continued), 3 (continued) ($m_n \leq n\Delta$).

8.1.2 Rates for the estimation on a compact set

In all the examples above, it is possible to find a compact set A such that g is of class C^∞ on A .

Due to Corollary 4.1, for all $\alpha > 0$, $\mathbb{E}(\|g - \tilde{g}_m\|_A^2) = O((n\Delta)^{-2\alpha/(2\alpha+1)})$. For the conditions under which this rate arises, three possibilities happen:

1. for the compound Poisson process with Gaussian and exponential density, we have $\int u^2 |g^*(u)|^2 du < +\infty$,
2. for the compound Poisson process with uniform density f , the Lévy Gamma process and the bilateral Lévy Gamma process, we have $\int u^2 |g^*(u)|^2 du = +\infty$ and g is bounded.
3. For the Lévy- δ process (see Example 2 (continued)), $\int u^2 |g^*(u)|^2 du = +\infty$ and g is not bounded.

Choosing $\Delta = n^{-a}$ (see Corollary 4.1), in the first case, the best rate corresponding to $\alpha \rightarrow +\infty$ is of order $O(n^{-2/3})$, for the second case, of order $O(n^{-2/5})$ and for the third case of order $O(n^{-1/3})$.

8.1.3 Comparison

To conclude, we give in Table 3 the best rate that can be obtained on each example according to the method, either Fourier method (with the Sinus Cardinal basis) or the time domain method (with the Trigonometric basis). The winner of the challenge is always the trigonometric basis. This is because the limit $\alpha \rightarrow +\infty$ is considered for the latter basis only. However, on simulations, the Fourier method performs better.

Process	Sinus Cardinal basis	Trigonometric basis
Poisson-Gaussian	$\log^{1/2}(n)n^{-2/3}$	$n^{-2/3}$
Poisson-Exp.	$n^{-1/2}$	$n^{-2/3}$
Poisson-Unif.	$n^{-1/4}$	$n^{-2/5}$
Lévy-Gamma	$n^{-1/4}$	$n^{-2/5}$
Lévy- δ	$n^{-\delta/(2\delta+1)}$, $\delta \in (0, 1/2)$	$n^{-1/3}$
Bilateral Gamma	$n^{-1/4}$	$n^{-2/5}$

Table 3 Comparison of best possible rates with the two methods.

In all cases, rates measured as powers of n are slower than in classical density estimation. The important value is $n\Delta$, that should be large enough. This means that Δ cannot be too small in order to keep a reasonable number n of observations.

8.2 General case

We consider the previous examples with the addition of a drift and a Brownian motion and look at the rates for the estimation of p deduced from Proposition 6.1 and Theorem 6.1. We indicate in which cases the estimation of σ is possible using the estimators $\hat{\sigma}(r)$.

Example 1. Drift + Brownian motion+ Compound poisson process.

Let

$$L_t = b_0t + \sigma W_t + \sum_{i=1}^{N_t} Y_i \quad (8.5)$$

where N_t is a Poisson process with constant intensity c and Y_i is a sequence of *i.i.d.* random variables with density f , independent of the process (N_t) .

Note that $\mathbb{E}L_1 = b = b_0 + \int xn(x)dx$ ($n(x) = cf(x)$). For the estimation of $p(x) = cx^3f(x)$, the rates that can be obtained depend on the density f provided that f satisfies the assumptions of Theorem 6.1, which are essentially here moment assumptions for the r.v.'s Y_i . Any order can be obtained as shown in Table 4 where rates are computed for f a standard Gaussian, an exponential with parameter 1 and a Beta distribution with parameters $(1, 3)$ (for p to be regular enough).

As $\int |x|^r n(x)dx < \infty$ for all $r < 1$ (actually, for all $r \leq 2$), estimation of σ is possible using $\hat{\sigma}(r)$ for any value of $0 < r < 1$ (provided that $n\Delta^{2-r} = o(1)$).

$f(x)$	$\mathcal{N}(0, 1)$	$\mathcal{E}(1)$	$\beta(1, 3)$
$p(x) = cx^3f(x)$	$\propto x^3 e^{-x^2}$	$\propto x^3 e^{-x} \mathbb{I}_{x>0}$	$\propto x^3(1-x)^2 \mathbb{I}_{[0,1]}(x)$
$p^*(u)$	$\propto (u^3 - 3u)e^{-u^2/2}$	$\propto 1/(1-iu)^4$	$O(1/ u ^3)$ for large $ u $.
$\int_{ u \geq \pi m} p^*(u) ^2 du$	$O((\pi m)^5 e^{-(\pi m)^2})$	$O((\pi m)^{-7})$	$O((\pi m)^{-5})$
$\int_{ u \leq \pi \mu_n} u^4 p^*(u) ^2 du$	$O(1)$	$O(1)$	$O(1)$
\check{m} (best choice of m)	$\sqrt{\log(n\Delta) - \frac{5}{2} \log \log(n\Delta)}/\pi$	$O((n\Delta)^{1/8})$	$O((n\Delta)^{1/6})$
Rate \propto	$\frac{\sqrt{\log(n\Delta)}}{n\Delta}$	$(n\Delta)^{-7/8}$	$(n\Delta)^{-5/6}$

Table 4 Rates for different "Drift+ Brownian motion +Compound Poisson processes" ($\mu_n \leq \sqrt{n\Delta}$).

Example 2. Drift + Brownian motion + Lévy-Gamma process.

Consider $L_t = b_0t + \sigma W_t + \Gamma_t$ where (Γ_t) is a Lévy gamma process with parameters (β, α) . We have $\mathbb{E}L_1 = b = b_0 + \int xn(x)dx$ and $p(x) = \beta x^2 e^{-\alpha x} \mathbb{I}_{x>0}$. Elementary computations show (with $\mu_n \leq \sqrt{n\Delta}$):

$$p^*(u) = 2\beta/(\alpha - iu)^3, \quad \int_{|u|\geq \pi m} |p^*(u)|^2 du = O(m^{-5}), \quad \int_{-\pi \mu_n}^{\pi \mu_n} u^4 |p^*(u)|^2 du = O(1).$$

Therefore the rate for estimating p is $O((n\Delta)^{-5/6})$ for a choice $\check{m} = O((n\Delta)^{1/6})$.

As for all $r > 0$, $\int x^r n(x)dx < \infty$, σ may be estimated by $\hat{\sigma}(r)$ for any value of $0 < r < 1$.

Example 2 (continued). Drift + Brownian motion + A specific class of subordinators.

Let $L_t = b_0t + \sigma W_t + \Gamma_t$ where (Γ_t) is a subordinator of pure jump type with Lévy

density of the form $n(x) = \beta x^{\delta-1/2} x^{-1} e^{-\alpha x} \mathbb{1}_{x>0}$ with $\delta > -1/2$ (thus $\int xn(x)dx < \infty$). This class of subordinators includes compound Poisson processes ($\delta > 1/2$) and Lévy Gamma processes ($\delta = 1/2$). When $\delta > 0$, the function $xn(x)$ is both integrable and square integrable (see above). Here, we can also consider the estimation of p in the case the case $-1/2 < \delta \leq 0$ which includes the Lévy Inverse Gaussian process ($\delta = 0$). The function $p(x) = x^3 n(x)$ can be estimated in presence (or not) of additional drift and Brownian component. We obtain:

$$p^*(u) = \beta \frac{\Gamma(\delta + 5/2)}{(\alpha - iu)^{\delta+5/2}} \quad \text{and} \quad \int_{|u| \geq \pi m} |p^*(u)|^2 du = O(m^{-(2\delta+4)}).$$

In the case $\delta \leq 0$, $u^4 |p^*(u)|^2$ is not integrable and we have for $n\Delta^2 \leq 1$,

$$\Delta^2 \int_{|u| \leq \pi \mu_n} u^4 |p^*(u)|^2 du = \Delta^2 o(\mu_n) = o(\Delta^{3/2}).$$

The best rate for estimating p is $O((n\Delta)^{-(2\delta+4)/(2\delta+5)})$ for a choice $\check{m} = O((n\Delta)^{1/(2\delta+5)})$. Note that $\Delta^{3/2} \leq (n\Delta)^{-(2\delta+4)/(2\delta+5)}$ for $n\Delta^2 \leq 1$ and $-1/2 < \delta \leq 0$.

For $r > 1/2 - \delta$, $\int x^r n(x)dx < \infty$. Hence, to estimate σ using $\hat{\sigma}(r)$, we must choose $1/2 - \delta < r < 1$.

Example 3. Drift + Brownian motion + Pure jump martingale.

Consider $L_t = bt + \sigma W_t + B_{\Gamma_t}$ where W, B, Γ are independent processes, W, B are standard Brownian motions, and Γ is a pure-jump subordinator with Lévy density $n_{\Gamma}(\gamma) = \beta \gamma^{\delta-1/2} \gamma^{-1} e^{-\alpha \gamma} \mathbb{1}_{\gamma>0}$ as above (assuming $\delta > -1$). The Lévy density $n(\cdot)$ of (L_t) (and of $(X_t = B_{\Gamma_t})$) is linked with n_{Γ} (see (3.3)) and can be computed:

$$n(x) = \frac{2\beta}{\sqrt{2\pi}} K_{\delta-1}(\sqrt{2\alpha}|x|) \left(\frac{|x|}{\sqrt{2\alpha}}\right)^{\delta-1},$$

where K_{ν} is a Bessel function of third kind (MacDonald function) (see e.g. [3]). For $\delta = 1/2$, B_{Γ_t} is a symmetric bilateral Lévy Gamma process. For $\delta = 0$, B_{Γ_t} is a normal inverse Gaussian Lévy process. The relation (3.3) allows to check that the function $p(x) = x^3 n(x)$ belongs to $\mathbb{L}^1 \cap \mathbb{L}^2$ and satisfies (H4-p) for $\delta > -3/4$. Moreover:

$$p^*(u) = -i\beta \left(\frac{u^3 \Gamma(\delta + 5/2)}{(\alpha + u^2/2)^{5/2}} - 3 \frac{u \Gamma(\delta + 3/2)}{(\alpha + u^2/2)^{3/2}} \right).$$

Thus, with $n\Delta^2 \leq 1$,

$$\int_{|u| \geq \pi m} |p^*(u)|^2 du = O(m^{-3}) \quad \text{and} \quad \Delta^2 \int_{|u| \leq \pi \mu_n} u^4 |p^*(u)|^2 du = \Delta^2 O(\mu_n) = O(\Delta^{3/2}).$$

The best rate for estimating p is $O((n\Delta)^{-3/4})$ obtained for $\check{m} = O((n\Delta)^{1/4})$. We have $\Delta^{3/2} \leq (n\Delta)^{-3/4}$ as $n\Delta^2 \leq 1$. As $\int \gamma^{r/2} n_{\Gamma}(\gamma) d\gamma < \infty$ for $r > 1 - \delta/2$, the estimation of σ by $\hat{\sigma}(r)$ requires $1 - \delta/2 < r < 1$. Therefore, we must have $\delta > 0$.

9 Simulations

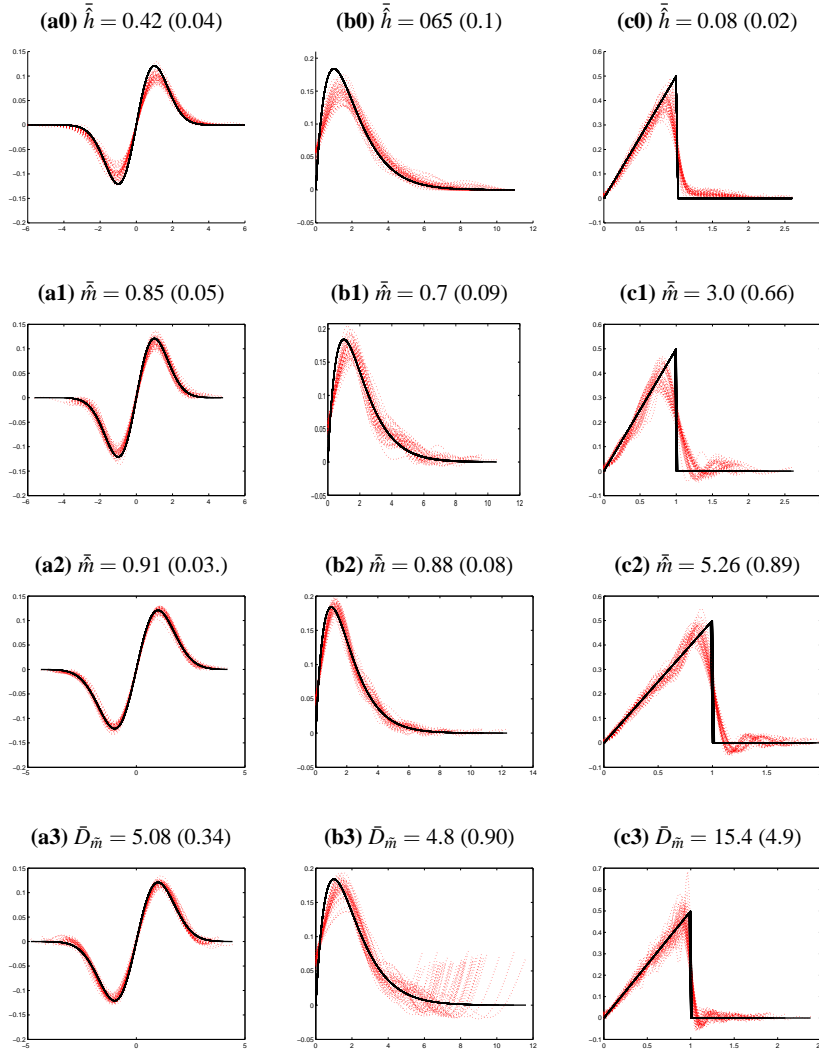


Fig. 1 Estimation of g for a compound Poisson process with Gaussian (first column), Exponential $\mathcal{E}(1)$ (second column), and uniform $\mathcal{U}([0, 1])$ (third column) Y_i 's, $c = 0.5$. True (bold black line) and 50 estimated curves (dotted red), left $\Delta = 0.2$ $n = 5000$: Kernel estimator (first line), Sinus Cardinal basis (second line); $\Delta = 0.05$, $n = 5 \cdot 10^4$: Sinus Cardinal basis (third line), trigonometric basis (fourth line).

In this section, we illustrate on numerical simulated data the performance of the estimators.

9.1 Simulations in the pure jump case

The adaptive estimation methods of Section 4 were implemented in the three cases: kernel method, deconvolution (Sinus Cardinal basis) and estimation of g on a compact subset using trigonometric bases. Lévy processes chosen among the examples given in Section 8.1 were simulated. Precisely,

1. A compound Poisson process with Gaussian $\mathcal{N}(0, 1)$ Y_i 's, $g(x) = cx \exp(-x^2/2)/\sqrt{2\pi}$.
2. A compound Poisson process with Exponential $\mathcal{E}(1)$ Y_i 's, $g(x) = cxe^{-x}\mathbb{I}_{x>0}$.
3. A compound Poisson process with Uniform $\mathcal{U}([0, 1])$ Y_i 's, $g(x) = cx\mathbb{I}_{[0,1]}(x)$.
4. A Lévy-Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$, $g(x) = \beta \exp(-\alpha x)\mathbb{I}_{x>0}$,
5. A Lévy-Gamma process with parameters $(\alpha, \beta) = (1, 1)$,
6. A Bilateral Lévy-Gamma process with parameters $(\alpha, \beta) = (\alpha', \beta') = (2, 0.2)$,
 $g(x) = \beta \exp(-\alpha x)\mathbb{I}_{x>0} - \beta' \exp(\alpha' x)\mathbb{I}_{x<0}$,
7. A Bilateral Lévy-Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$ and $(\alpha', \beta') = (1, 1)$

The implementation of the adaptive method requires the calibration of the constant κ in the penalties. This is a difficulty of the method. In practice, the penalty constant is usually calibrated by preliminary simulations. After this was done, the constant κ was taken equal to 1.5 in the kernel method, to 7.5 for the deconvolution and to 1 when using the trigonometric basis. The bandwidth \hat{h} was chosen among 20 equispaced values between 0.01 and 0.75 with a standard Gaussian kernel, to ease the computation of the iterated kernel $K_h \star K_{h'}$. The cut-off \hat{m} was chosen among 100 equispaced values between 0 and 10. The dimension $D_{\hat{m}}$ was chosen among 80 values between 1 and 80. We used in both cases the expression of the estimators using their coefficients on the bases. In the Sinus Cardinal case, this avoids high dimensional matrices manipulations, but the series have to be truncated (we kept coefficients $\hat{a}_{m,j}$ for $|j| \leq K_n$ with $K_n = 15$).

Results are given in Figures 1 and 2 where 50 estimated curves are plotted on the same figure to illustrate the weak variability of the estimator. In Figure 1, estimation results for compound Poisson processes are plotted. The first line illustrates the kernel method, the second and third lines give estimation results with the Sinus Cardinal basis and the fourth line concerns the trigonometric basis. In the first two lines, we choose $n = 5000, \Delta = 0.2$ ($n\Delta = 1000$) and in the last two lines, $n = 50000, \Delta = 0.05$ ($n\Delta = 2500$).

Figure 2 illustrates the estimation of Lévy Gamma models. In the first two columns, curves are estimated by Sinus Cardinal basis, while the last columns concerns the trigonometric basis.

It is clear from both Figures 1 and 2 that increasing $n\Delta$ improves the result by showing a thinner variability band. Comparing the last two lines of Figure 1 and

the last two columns of Figure 2 amounts to comparing the performance of the two bases. It appears that the Sinus Cardinal must be preferred because the trigonometric basis has very important edge effects for highly dissymmetric densities: see in particular the exponential-Poisson, and the Gamma case, which start with a peak and end at zero. The kernel and the deconvolution methods seem to have analogous performances.

On top of each graph in Figures 1 and 2, the mean of the selected values for \hat{h} , \hat{m} (sinus cardinal basis) or for $D_{\hat{m}}$ (trigonometric basis) is given with the associated standard deviation in parentheses. Various values are chosen by the estimation procedure, and in each case, the standard deviation exhibits a reasonable variability. This is an indication that the constants in the penalties are adequately chosen: too small constants κ imply very unstable choices for the same model, while greater κ 's quickly lead to null standard deviations for 50 sample paths. Note also that the higher the regularity of g , the smaller the selected $1/\hat{h}$'s, \hat{m} 's and $D_{\hat{m}}$'s (which is coherent with orders $O(n^{1/(2\alpha+1)})$ for a regularity α). The uniform-Poisson case involves larger values for $1/\hat{h}$, \hat{m} , $D_{\hat{m}}$ than the two other Poisson cases, for instance.

9.2 General case and comparisons

In this section, we present numerical results for simulated Lévy processes corresponding to Examples 1 and 2 of Section 8.2. For these models, the functions $g(x) = xn(x)$, ℓ and p belong to $\mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R})$. Thus, we can estimate g when $b_0 = 0$, $\sigma = 0$, ℓ when $\sigma = 0$ and p when $\sigma \neq 0$. The estimators $\hat{g}_{\hat{m}}$, $\hat{\ell}_{\hat{m}}$, $\hat{p}_{\hat{m}}$ using the sinus cardinal basis were implemented (see (5.6)-(5.14) and (6.3)-(6.6)). After preliminary experiments, the numerical constants κ , κ' appearing in the penalties were set to 7.5 for g , 4 for ℓ and 3 for p . The cut-off \hat{m} was chosen among 100 equispaced values between 0 and 10.

Figure 3 shows estimated curves for models with jump part coming from compound Poisson processes (see (8.5)) where the Y_i 's are standard Gaussian, Exponential $\mathcal{E}(1)$, and $\beta(3, 3)$ rescaled on $[-4, 4]$. The intensity c is equal to 0.5.

Figure 4 shows estimated curves for jump part of Lévy Gamma and bilateral Lévy Gamma type. The bilateral Lévy Gamma process is the difference $\Gamma_t - \Gamma_t'$ of two independent Lévy Gamma processes.

On top of each graph, we give the mean value of the selected cut-off with its standard deviation in parentheses. This value is surprisingly small. As expected, the presence of a Gaussian component deteriorates the estimation, which remains satisfactory on the whole.

Generally, authors estimate $n(\cdot)$ on a compact set separated from the origin (see e.g. [31]). Setting $\hat{n}(x) = \hat{g}(x)/x$, we have the obvious inequality

$$\mathbb{E}(\|(\hat{n} - n)1_{\mathbb{R}/[-a, a]}\|^2) \leq \frac{1}{a^2} \mathbb{E}(\|\hat{g} - g\|^2).$$

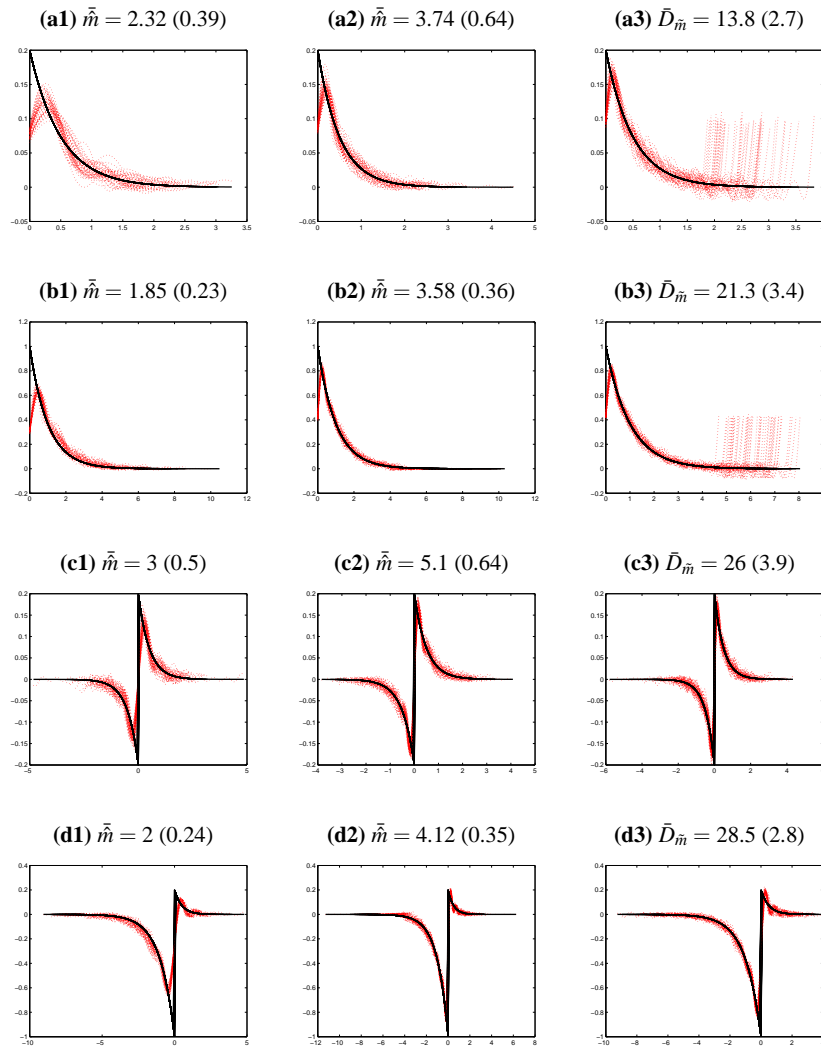


Fig. 2 Estimation of g for a Lévy Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$ (first line), $(\alpha, \beta) = (1, 1)$ (second line), a bilateral Lévy Gamma process with parameters $(\alpha, \beta) = (\alpha', \beta') = (2, 0.2)$ (third line) and a bilateral Lévy Gamma process with parameters $(\alpha, \beta) = (2, 0.2)$, $(\alpha', \beta') = (1, 1)$. True (bold black line) and 50 estimated curves (dotted red), left $\Delta = 0.2$, $n = 5000$, Sinus Cardinal basis; center, $\Delta = 0.05$, $n = 5 \cdot 10^4$, Sinus Cardinal basis; right $\Delta = 0.05$, $n = 5 \cdot 10^4$, trigonometric basis.

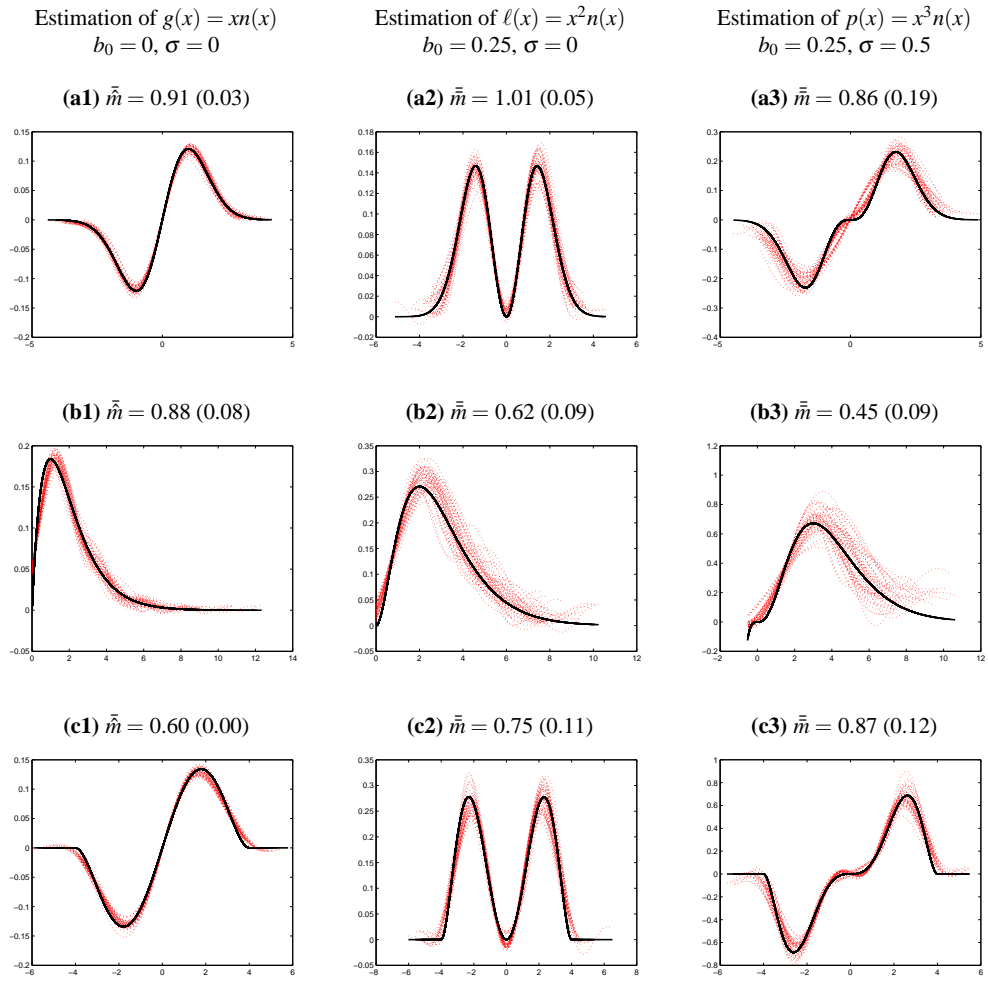


Fig. 3 Variability bands for the estimation of g, ℓ, p for a compound Poisson process with Gaussian (first line), Exponential $\mathcal{E}(1)$ (second line) and $\beta(3, 3)$ rescaled on $[-4, 4]$ (third line) Y_i 's, with $c = 0.5$. True (bold black line) and 50 estimated curves (dotted red), $\Delta = 0.05, n = 5 \cdot 10^4$.

Analogous inequalities hold for $\hat{n}(x) = \bar{\ell}(x)/x^2$ or $\hat{n}(x) = \bar{p}(x)/x^3$. In Figure 5, the estimator of $n(\cdot)$ deduced by dividing by the correct power of x is plotted, excluding an interval $[-a, a]$ around zero. To obtain correct representations, $a = 0.1$ suits for $\hat{g}(x)/x$, $a = 0.5$ for $\bar{\ell}(x)/x^2$ and $a = 1$ for $\bar{p}(x)/x^3$. The results are satisfactory and in accordance with the difficulty of estimating $n(\cdot)$ without or with Gaussian component.

Tables 5 and 6 show the means of the estimation results for $b = \mathbb{E}(L_1) = b_0 + \int xn(x)dx$ (see (7.1)) and σ , with standard deviations in parentheses.

The estimation of b is good in all cases, and especially when $n\Delta$ is large. The estimation of σ is clearly more difficult, with noticeable differences according to the values of n and Δ . When Δ is not small enough, the estimation can be heavily biased. In accordance with the theory, when r is smaller, the estimator of σ is slightly better (smaller bias). Table 7 shows the values of $n\Delta^2$ and $n\Delta^{2-r}$, which should be small for the performance of the estimator to be satisfactory. It is worth noting that σ is constantly over-estimated.

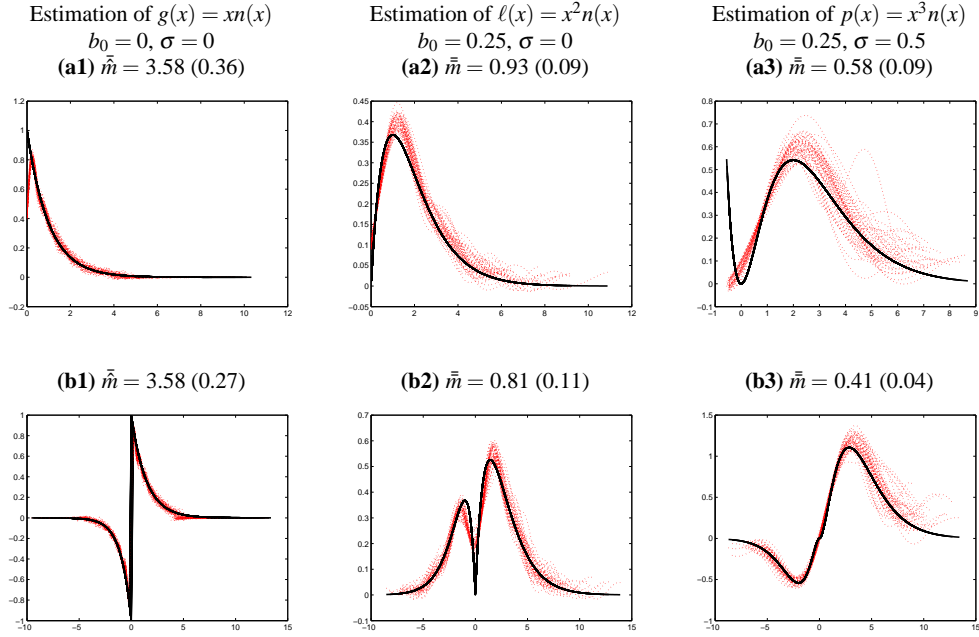


Fig. 4 Variability bands for the estimation of g, ℓ, p for jumps from a Lévy-Gamma process with $\beta = 1, \alpha = 1$ (first line), a bilateral Lévy-Gamma process with $(\beta, \alpha) = (0.7, 1), (\beta', \alpha') = (1, 1)$ (second line). True (bold black line) and 50 estimated curves (dotted red), $\Delta = 0.05, n = 5.10^4$.

10 Compound Poisson processes

This section is devoted to compound Poisson processes which are a special case of Lévy processes with integrable Lévy measure. Compound Poisson processes are widely used in practice especially in queuing and insurance theory (see *e.g.* [27] and references therein, [47] or [59]). The results given here are based on the paper [16]. One advantage of the approach is to weaken the constraints on the sampling interval. Let $(X_t, t \geq 0)$ be a compound Poisson process, given by

Model	(n, Δ)	$(5 \cdot 10^4, 0.05)$	$(5 \cdot 10^4, 0.01)$	$(5 \cdot 10^4, 10^{-3})$	$(10^4, 10^{-3})$
Poisson	$\hat{b} (b = 1)$	1.000 (0.02)	0.997 (0.04)	0.995 (0.123)	1.001 (0.280)
Gaussian	$\hat{\sigma}(1/2)$	0.602 (0.03)	0.527 (0.002)	0.504 (0.002)	0.504 (0.005)
	$\hat{\sigma}(1/4)$	0.589 (0.03)	0.521 (0.002)	0.503 (0.002)	0.503 (0.002)
Poisson	$\hat{b} (b = 1.5)$	1.502 (0.05)	1.502 (0.051)	1.494 (0.142)	1.461 (0.359)
Exp(1)	$\hat{\sigma}(1/2)$	0.611 (0.003)	0.530 (0.003)	0.505 (0.002)	0.505 (0.005)
	$\hat{\sigma}(1/4)$	0.594 (0.003)	0.522 (0.003)	0.503 (0.002)	0.503 (0.005)
Gamma	$\hat{b} (b = 2)$	2.001 (0.02)	2.000 (0.05)	1.998 (0.177)	2.018 (0.335)
(1,1)	$\hat{\sigma}(1/2)$	0.705 (0.004)	0.562 (0.003)	0.512 (0.002)	0.513 (0.005)
	$\hat{\sigma}(1/4)$	0.677 (0.004)	0.548 (0.003)	0.508 (0.002)	0.508 (0.005)
Bilateral	$\hat{b} (b = 1.4286)$	1.426 (0.035)	1.4286 (0.076)	1.4493 (0.264)	1.405 (0.619)
Gamma	$\hat{\sigma}(1/2)$	0.862 (0.005)	0.628 (0.004)	0.526 (0.003)	0.526 (0.006)
(0.7,1), (1,1)	$\hat{\sigma}(1/4)$	0.798 (0.004)	0.593 (0.003)	0.516 (0.002)	0.515 (0.006)

Table 5 Estimation of (b, σ) , $b_0 = 1$, the true value of b in parenthesis, $\sigma = 0.5$, $K = 200$ replications.

Model	(n, Δ)	$(5 \cdot 10^4, 0.05)$	$(5 \cdot 10^4, 0.01)$	$(5 \cdot 10^4, 10^{-3})$	$(10^4, 10^{-3})$
Poisson	$\hat{b} (1)$	0.999 (0.025)	1.005 (0.059)	0.998 (0.178)	1.025 (0.85)
Gaussian	$\hat{\sigma}(1/2)$	1.082 (0.005)	1.026 (0.004)	1.006 (0.004)	1.005 (0.009)
	$\hat{\sigma}(1/4)$	1.072 (0.005)	1.020 (0.005)	1.004 (0.004)	1.003 (0.01)
Poisson	$\hat{b} (1.5)$	1.510 (0.026)	1.498 (0.06)	1.481 (0.190)	1.485 (0.442)
Exp(1)	$\hat{\sigma}(1/2)$	1.096 (0.005)	1.030 (0.004)	1.006 (0.004)	1.006 (0.009)
	$\hat{\sigma}(1/4)$	1.080 (0.005)	1.022 (0.004)	1.003 (0.004)	1.003 (0.010)
Gamma	$\hat{b} (2)$	2.00 (0.026)	1.995 (0.068)	1.991 (0.196)	2.023 (0.195)
(1,1)	$\hat{\sigma}(1/2)$	1.172 (0.005)	1.062 (0.005)	1.014 (0.004)	1.014 (0.004)
	$\hat{\sigma}(1/4)$	1.152 (0.005)	1.050 (0.005)	1.010 (0.005)	1.010 (0.004)
Bilateral	$\hat{b} (1.4286)$	1.425 (0.04)	1.431 (0.10)	1.429 (0.28)	1.492 (0.63)
Gamma	$\hat{\sigma}(1/2)$	1.330 (0.006)	1.136 (0.005)	1.033 (0.005)	1.033 (0.01)
(0.7,1), (1,1)	$\hat{\sigma}(1/4)$	1.284 (0.006)	1.105 (0.005)	1.022 (0.005)	1.022 (0.01)

Table 6 Estimation of (b, σ) , $b_0 = 1$, the true value of b in parenthesis, $\sigma = 1$, power variation method for estimation of σ , $K = 200$ replications.

(n, Δ)	$(5 \cdot 10^4, 0.05)$	$(5 \cdot 10^4, 0.01)$	$(5 \cdot 10^4, 10^{-3})$	$(10^4, 10^{-3})$
$n\Delta$	2500	500	50	10
$n\Delta^2$	125	5	0.05	0.01
$n\Delta^{2-1/2}$	559	50	1.6	0.3
$n\Delta^{2-1/4}$	264	16	0.3	0.06

Table 7 Values of $n, \Delta, n\Delta, n\Delta^2, n\Delta^{2-r}$ for $r = 1/2$ and $r = 1/4$.

$$X_t = \sum_{i=1}^{N_t} \xi_j, \quad (10.1)$$

where $(\xi_j, j \geq 1)$ is a sequence of *i.i.d.* real valued random variables with density f , (N_t) is a Poisson process with intensity $c > 0$, independent of the sequence $(\xi_j, j \geq 1)$. The density f and the intensity c are unknown. We are interested in

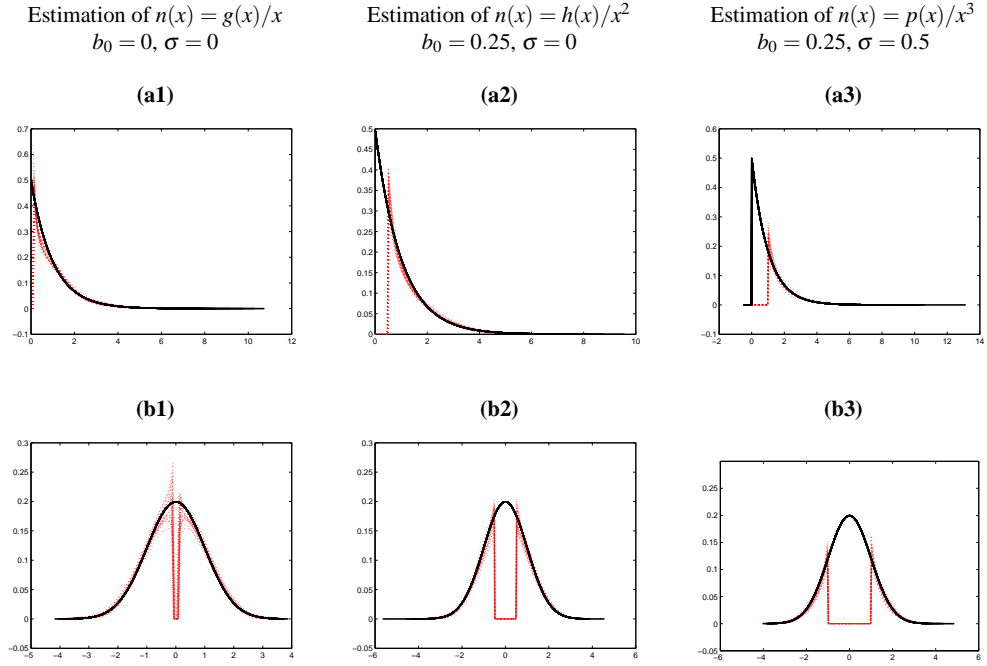


Fig. 5 Variability bands for the estimation of $n(\cdot)\mathbb{I}_{[-a,a]^c}$ for a compound Poisson process with Exponential $\mathcal{E}(1)$ (first line) and Gaussian (second line) jump densities, with $a = 0.1$ (first column), $a = 0.5$ (second column), $a = 1$ (third column). In all cases, $c = 0.5$, $n = 50000$, $\Delta = 0.05$; 25 estimated curves (thin dotted) and the true (bold line).

adaptive nonparametric estimation of f from discrete observations $(X_{j\Delta}, j \geq 0)$ and the resulting estimation of the Lévy density $n(x) = cf(x)$ where the intensity c has to be estimated too. As compound Poisson processes are simpler than general Lévy processes, specific methods for estimating the jump distribution have been investigated. The estimation of f is often called decompounding (see for instance, [12], [28] or [25]). We adopt the point of view of [28] to define the discrete observations of the sample path (X_t) .

Recall that the common distribution of the increments $X_{k\Delta} - X_{(k-1)\Delta}$ is equal to

$$\mathbb{P}_{X_\Delta}(dx) = e^{-c\Delta} \delta_0(dx) + (1 - e^{-c\Delta}) q_\Delta(x) dx, \quad (10.2)$$

where δ_0 is the Dirac mass at 0, q_Δ is the conditional density of X_Δ given that $X_\Delta \neq 0$:

$$q_\Delta = \sum_{m \geq 1} \frac{e^{-c\Delta}}{1 - e^{-c\Delta}} \frac{(c\Delta)^m}{m!} f^{*m}, \quad (10.3)$$

and f^{*m} denotes the m -th convolution power of f . As null increments provide no information on the density f , we assume that the sample path X_t is discretely ob-

served until exactly n increments are nonzero. Such observations can be described as follows. Let

$$\begin{aligned} S_1 &= \inf\{j \geq 1, X_{j\Delta} - X_{(j-1)\Delta} \neq 0\}, \\ S_i &= \inf\{j > S_{i-1}, X_{j\Delta} - X_{(j-1)\Delta} \neq 0\}, i \geq 2, \end{aligned} \quad (10.4)$$

and set

$$Z_i = X_{S_i\Delta} - X_{(S_i-1)\Delta}. \quad (10.5)$$

(For the sake of simplicity, in this section, we use the same notation Z_i for the above increments). Assume that the $X_{j\Delta}$'s are observed for $j \leq S_n$. Thus, $(S_i, Z_i), i = 1, \dots, n$ are observed. Proposition 10.1 gives the joint distribution of these observations. In particular, it is shown that Z_1, \dots, Z_n is a n -sample of the conditional distribution of X_Δ given that $X_\Delta \neq 0$ which has density q_Δ . Therefore, the estimation of q_Δ is possible using the sample Z_1, \dots, Z_n . On the other hand, estimators of c can be based on (S_1, \dots, S_n) .

We use the following method to build an estimator of f . The operator $f \rightarrow q_\Delta := P_\Delta f$ can be explicitly inverted. Provided that $c\Delta < \log 2$, the inverse operator P_Δ^{-1} admits a series development implying that:

$$f = P_\Delta^{-1}(g_\Delta) = \sum_{m \geq 1} \frac{(-1)^{m+1}}{m} \frac{(e^{c\Delta} - 1)^m}{c\Delta} q_\Delta^{*m}. \quad (10.6)$$

Consequently, truncating the above development and keeping $K + 1$ terms, f can be approximated:

$$f \simeq \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{(e^{c\Delta} - 1)^m}{c\Delta} q_\Delta^{*m}. \quad (10.7)$$

The approximation is valid for small Δ . To estimate f , we replace, for $m = 1, \dots, K + 1$, $(e^{c\Delta} - 1)^m / c\Delta$ by adequate estimators and each q_Δ^{*m} by a nonparametric estimator based on the observations $(Z_j, j = 1, \dots, n)$ given by (10.5). The interest of the method is that, from the n -sample of the density q_Δ , \sqrt{n} -consistent nonparametric estimators of the convolution power q_Δ^{*m} , for $m \geq 2$, can be built (see e.g. [60]). Here, we adopt the method described in [14]. Of course, $m \geq 2$ is fixed and should not be too large. To simplify notations, we omit the dependence on Δ for q_Δ and set

$$q := q_\Delta, \quad q^{*m} := q_\Delta^{*m}. \quad (10.8)$$

First, we deal with the parametric estimation of c and the coefficients $c_m(\Delta)$. Second, the estimation of q^{*m} is described. Finally, the estimators of f and cf are given.

10.1 Parameter estimation

This section concerns the estimation of c and the coefficients $c_m(\Delta)$, $m \geq 1$ appearing in the series development (10.6) of f . This relies on the joint distribution of $S_i, Z_i, i \geq 1$.

Proposition 10.1 *Let $S_0 = 0$ and $S_i, Z_i, i \geq 1$ be given by (10.4)-(10.5). We have, for all $i \geq 1$, $\mathbb{P}(S_i < +\infty) = 1$, $(S_i - S_{i-1}, Z_i), i \geq 1$ are independent and identically distributed random couples. For $k \geq 1$,*

$$\mathbb{P}(S_1 = k, Z_1 \leq x) = e^{-c(k-1)\Delta}(1 - e^{-c\Delta})\mathbb{P}(X_\Delta \leq x | X_\Delta \neq 0).$$

Consequently, S_1 and Z_1 are independent, the distribution of Z_1 is equal to the conditional distribution of X_Δ given $X_\Delta \neq 0$, S_1 has geometric distribution with parameter $1 - e^{-c\Delta}$. Moreover, the random variables $(S_1, Z_1, \dots, S_i - S_{i-1}, Z_i, \dots, S_n - S_{n-1}, Z_n)$ are independent.

Proof. To obtain the joint distribution of (S_1, Z_1) is elementary using that the increments $X_{j\Delta} - X_{(j-1)\Delta}$ are *i.i.d.*. The process $(X_{j\Delta}^x = x + X_{j\Delta}, j \geq 1)$ is strong Markov. We denote by \mathbb{P}_x its distribution on the canonical space $\mathbb{R}^{\mathbb{N}}$, denote by $(X_j, j \geq 0)$ the canonical process of $\mathbb{R}^{\mathbb{N}}$ and by $\mathcal{F}_j = \sigma(X_k, k \leq j)$ the canonical filtration. Let $\theta : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}^{\mathbb{N}}$ denote the shift operator. Consider the stopping times built on the canonical process $S_0 = 0$,

$$S_i = \inf\{j > S_{i-1}, X_j - X_{j-1} \neq 0\}, i \geq 1,$$

and let

$$Z_i = X_{S_i} - X_{S_{i-1}}.$$

Because the S_i 's are built using the increments $(X_j - X_{j-1}, j \geq 1)$, their distributions under \mathbb{P}_x are independent of the initial condition x . We have $S_i = S_{i-1} + S_1 \circ \theta_{S_{i-1}}$. The process $(X_{S_{i-1}+j} - X_{S_{i-1}} = (X_j - X_0) \circ \theta_{S_{i-1}}, j \geq 0)$ is independent of $\mathcal{F}_{S_{i-1}}$ and has distribution \mathbb{P}_0 and $Z_i = Z_1 \circ \theta_{S_{i-1}}$. Consequently,

$$\mathbb{E}_x(\varphi(S_i - S_{i-1})\psi(Z_i) | \mathcal{F}_{S_{i-1}}) = \mathbb{E}_0(\varphi(S_1)\psi(Z_1)).$$

By iterate conditioning, we get the result.

Let us now turn to the estimation of $c_m(\Delta)$ for all $m \geq 1$ and c . For this, we use the sample (S_1, \dots, S_n) which is independent of the sample (Z_1, \dots, Z_n) . As we deal with a semiparametric problem, we need find estimators with computable \mathbb{L}^2 -risk. So the simple plug-in of the exact maximum likelihood estimator of c is not suitable.

Proposition 10.2 *Assume that $c \in [c_0, c_1]$ with $c_0 > 0$ and $c_1\Delta \leq \log(2)/2$. Let*

$$F(\xi) = \frac{1}{\Delta} \log \frac{\xi}{\xi - 1} \tag{10.9}$$

and for $m \geq 1$

$$H_m(\xi) = \frac{1}{(\xi - 1)^m \log \frac{\xi}{\xi - 1}}. \quad (10.10)$$

Define

$$\begin{aligned} \Omega_n &= \left\{ 1 + \frac{1}{e^{2c_1\Delta} - 1} \leq \frac{S_n}{n} \leq 1 + \frac{1}{e^{c_0/(2\Delta)} - 1} \right\}, \\ \widehat{c_m(\Delta)} &= H_m(S_n/n) 1_{\Omega_n}, \quad \widehat{c} = F(S_n/n) 1_{\Omega_n}. \end{aligned} \quad (10.11)$$

Then,

$$\mathbb{E} \left(\widehat{c_m(\Delta)} - c_m(\Delta) \right)^2 \leq C_m \frac{\Delta^{2(m-1)}}{n}, \quad \mathbb{E} (\widehat{c} - c)^2 \leq \frac{C}{n}, \quad (10.12)$$

where C_m, C have an explicit expression as functions of c_0, c_1 and m .

Note that the bounds are non asymptotic and the exact value of the constants C_m, C can be deduced from the proof.

Proof. We start with the estimators of $c_m(\Delta)$. Let us set

$$p(\Delta) = 1 - e^{-c\Delta} = \frac{e^{c\Delta} - 1}{e^{c\Delta}}.$$

An elementary computation yields:

$$c\Delta = \log\left(\frac{x}{x-1}\right) \quad \text{with} \quad x := x(\Delta) = \frac{1}{p(\Delta)} = 1 + \frac{1}{e^{c\Delta} - 1} > 1,$$

and

$$\frac{(e^{c\Delta} - 1)^m}{c\Delta} = H_m(x).$$

As the standard maximum likelihood (and unbiased) estimator of $1/p(\Delta)$ computed from the sample $(S_i - S_{i-1}, i = 1, \dots, n)$ is $S_n/n \geq 1$, we are tempted to estimate $H_m(x)$ by $H_m(S_n/n)$. This is not possible as S_n/n may be equal to 1. This is why we introduce a truncation. Set $u_0 = \Delta/(e^{c_0\Delta/2} - 1)$, $u_1 = \Delta/(e^{2c_1\Delta} - 1)$, $u = \Delta/(e^{c\Delta} - 1)$. Note that

$$1 + \frac{u_1}{\Delta} < x = 1 + \frac{u}{\Delta} < 1 + \frac{u_0}{\Delta}, \quad \Omega_n = \left\{ 1 + \frac{u_1}{\Delta} \leq \frac{S_n}{n} \leq 1 + \frac{u_0}{\Delta} \right\}. \quad (10.13)$$

We have

$$\widehat{c_m(\Delta)} - c_m(\Delta) = H_m(S_n/n) 1_{\Omega_n} - H_m(x) = A_1 + A_2$$

with

$$A_1 = (H_m(S_n/n) - H_m(x)) 1_{\Omega_n},$$

and

$$A_2 = -H_m(x) 1_{\Omega_n^c}.$$

On Ω_n ,

$$(H_m(S_n/n) - H_m(x))^2 \leq \left(\frac{S_n}{n} - x\right)^2 \sup_{\xi \in [1 + \frac{u_1}{\Delta}, 1 + \frac{u_0}{\Delta}]} (H'_m(\xi))^2.$$

As

$$H'_m(\xi) = -\frac{m}{(\xi - 1)^{m+1} \log \frac{\xi}{\xi-1}} + \frac{1}{\xi(\xi - 1)^{m+1} \log^2 \frac{\xi}{\xi-1}},$$

we have, for $\xi \in [1 + \frac{u_1}{\Delta}, 1 + \frac{u_0}{\Delta}]$,

$$|H'_m(\xi)| \leq \frac{2\Delta^m}{c_0 u_1^{m+1}} \left(m + \frac{2}{u_1 c_0}\right).$$

Writing that $e^{2c_1\Delta} - 1 = 2c_1\Delta e^{2sc_1\Delta}$ for $s \in (0, 1)$ and using that $2c_1\Delta \leq \log(2)$, we get $1/u_1 \leq 4c_1$. As

$$\mathbb{E}\left(\frac{S_n}{n} - x\right)^2 = \frac{1 - p(\Delta)}{np^2(\Delta)} = \frac{e^{c\Delta}}{n(e^{c\Delta} - 1)^2} (\sim \frac{1}{n\Delta^2}),$$

we obtain, using $e^{c\Delta} - 1 \geq c\Delta \geq c_0\Delta$:

$$\mathbb{E}A_1^2 \leq C'_m \frac{\Delta^{2(m-1)}}{n}, \text{ with } C'_m = \frac{4\sqrt{2}(4c_1)^{2(m+1)}}{c_0^4} \left(m + \frac{8c_1}{c_0}\right)^2.$$

Then, we have, setting $a_0 = u_0 - u > 0$, $a_1 = u - u_1 > 0$,

$$\begin{aligned} \mathbb{P}(\Omega_n^c) &= \mathbb{P}\left(\frac{S_n}{n} < 1 + \frac{u_1}{\Delta}\right) + \mathbb{P}\left(\frac{S_n}{n} > 1 + \frac{u_0}{\Delta}\right) \\ &= \mathbb{P}\left(\frac{\Delta}{p(\Delta)} - \Delta \frac{S_n}{n} > a_1\right) + \mathbb{P}\left(\Delta \frac{S_n}{n} - \frac{\Delta}{p(\Delta)} > a_0\right) \\ &\leq \left(\frac{1}{a_1^2} + \frac{1}{a_0^2}\right) \frac{\Delta^2 e^{c\Delta}}{n(e^{c\Delta} - 1)^2} (\sim \frac{1}{n}). \end{aligned}$$

Thus, noting that $u_0 - u \geq 1/(2c_1)$ and $u - u_1 \geq 1/(4\sqrt{2}c_0)$,

$$\mathbb{E}A_2^2 \leq \left(\frac{1}{a_1^2} + \frac{1}{a_0^2}\right) \frac{(e^{c\Delta} - 1)^{2(m-1)} e^{c\Delta}}{nc^2} \leq C''_m \frac{\Delta^{2(m-1)}}{n}, \quad (10.14)$$

where

$$C''_m = 4\sqrt{2} [8c_0^2 + c_1^2] \frac{(4c_1)^{2(m-1)}}{c_0^2}.$$

The proof is complete with $C_m = 2(C'_m + C''_m)$.

We proceed analogously for studying \hat{c} . As $x = 1 + (e^{c\Delta} - 1)^{-1}$ and $c_0 \leq c \leq c_1$,

$$\sup_x F(x) = 2c_1.$$

The derivative $F'(x) = -(\Delta x(x-1))^{-1}$ satisfies,

$$\sup_x |F'(x)| = \frac{(e^{c_1\Delta} - 1)^2}{\Delta e^{2c_1\Delta}}.$$

Therefore,

$$(\widehat{c} - c)^2 \leq \left(\frac{S_n}{n} - x\right)^2 4c_1^2 \Delta^2 e^{4c_1\Delta} + 2c_1 \mathbf{1}_{\Omega_n^c}.$$

Thus,

$$\mathbb{E}(\widehat{c} - c)^2 \leq 16\sqrt{2} \frac{c_1^2}{nc_0^2} + 2c_1 \mathbb{P}(\Omega_n^c) = \frac{C}{n}.$$

10.2 Estimation of the m -th convolution power of a density from a n -sample

This paragraph relies on [14]. Consider an *i.i.d.* sample of variables Z_1, \dots, Z_n with density q and characteristic function q^* , the Fourier transform of q . As $(q^*)^m$ is the Fourier transform of q^{*m} , [14] propose to estimate $(q^*)^m$ for all $m \geq 1$, by its empirical counterpart $(\tilde{q}^*(t))^m$, where:

$$\tilde{q}^*(t) = \frac{1}{n} \sum_{j=1}^n e^{itZ_j}, \quad (10.15)$$

Fourier inversion leads to the estimator with cut-off d ,

$$\widehat{q_d^{*m}}(x) = \frac{1}{2\pi} \int_{-\pi d}^{\pi d} e^{-itx} (\tilde{q}^*(t))^m dt. \quad (10.16)$$

The following bounds hold.

Proposition 10.3 For $m \geq 2$ and all t ,

$$\mathbb{E}(|\widehat{(q^*)^m}(t) - (q^*)^m(t)|^2) \leq \mathcal{E}_m \left(\frac{1}{n^m} + \frac{|q^*(t)|^2}{n} \right) \quad (10.17)$$

where \mathcal{E}_m is a constant which does not depend on n nor on q , increasing with m and $\widehat{(q^*)^m}(t) = (\tilde{q}^*(t))^m$. Consequently,

$$\mathbb{E}(\|\widehat{q_d^{*m}} - q^{*m}\|^2) \leq \frac{1}{2\pi} \int_{|t| \geq \pi d} |(q^{*m})^*(t)|^2 dt + \mathcal{E}_m \left(\frac{d}{n^m} + \frac{\|q\|^2}{n} \right). \quad (10.18)$$

Proof. First we state a useful Lemma.

Lemma 10.1 Let $(u, v) \in \mathbb{C}^2$ such that $|u| \leq 1$ and $|v| \leq 1$. Then, for any integer $m \geq 1$, we have

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

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

Proof of Lemma 10.1. For $m = 1$, the desired inequality is obviously satisfied with $E_m = 0$. Let us now investigate the case $m \geq 2$. By the binomial formula

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

As $|u| \leq 1$ and $|v| \leq 1$,

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

with

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

Lemma 10.1 is proved. \square

It follows from the inequalities $|\tilde{q}^*(t)| \leq 1$, $|q^*(t)| \leq \|q\|_1 = 1$, Lemma 10.1 and the elementary inequality $(x + y)^2 \leq 2(x^2 + y^2)$, $(x, y) \in \mathbb{R}^2$, that

$$|(\tilde{q}^*(t))^m - (q^*(t))^m|^2 \leq 2 (|\tilde{q}^*(t) - q^*(t)|^{2m} + E_m^2 |\tilde{q}^*(t) - q^*(t)|^2 |q^*(t)|^2).$$

Then, the Rosenthal Inequality implies the existence of a constant $C_m > 0$ such that

$$\mathbb{E} (|\tilde{q}^*(t) - q^*(t)|^{2m}) \leq C_m / n^m.$$

This implies that

$$\mathbb{E} (|(\tilde{q}^*(t))^m - (q^*(t))^m|^2) \leq \mathcal{E}_m \left(\frac{1}{n^m} + \frac{1}{n} |q^*(t)|^2 \right). \quad (10.19)$$

This ends the proof of (10.17).

For the second inequality, setting

$$q_d^{*m}(x) = \frac{1}{2\pi} \int_{-\pi d}^{\pi d} (q^*(t))^m e^{itx} dt, \quad x \in \mathbb{R}, \quad (10.20)$$

we obtain the usual decomposition

$$\mathbb{E} \left(\|\widehat{q_d^{*m}} - q^{*m}\|^2 \right) \leq 2 (\|q_d^{*m} - q^{*m}\|^2 + \mathbb{E} \left(\|\widehat{q_d^{*m}} - q_d^{*m}\|^2 \right)). \quad (10.21)$$

with

$$\|q_d^{*m} - q^{*m}\|^2 = \frac{1}{2\pi} \int_{|t| \geq \pi d} |q^*(t)|^{2m} dt, \quad (10.22)$$

$$\mathbb{E} \left(\|\widehat{q_d^{*m}} - q_d^{*m}\|^2 \right) = \frac{1}{2\pi} \int_{-\pi d}^{\pi d} \mathbb{E} (|\bar{q}(t)|^m - (q^*(t))^m|^2) dt \quad (10.23)$$

and

$$\int_{-\pi d}^{\pi d} |q^*(t)|^2 dt \leq \|q^*\|_2^2 = 2\pi \|q\|_2^2 \leq C. \quad (10.24)$$

It follows from (10.21), (10.22) and (10.24) that

$$\mathbb{E} \left(\|\widehat{q_d^{*m}} - q_d^{*m}\|^2 \right) \leq C \left(\frac{d}{n^m} + \frac{1}{n} \int_{-\pi d}^{\pi d} |q^*(t)|^2 dt \right) \leq C \left(\frac{d}{n^m} + \frac{1}{n} \right). \quad (10.25)$$

Plugging (10.25) and (10.22) in (10.21) implies Inequality (10.18).

We can discuss now the rates of convergence implied by the above proposition. Let q^{*m} belongs to the Sobolev class $\mathcal{C}(a_m, R_m)$ (see (4.14)). The \mathbb{L}^2 -risk bound becomes

$$\mathbb{E}(\|\widehat{q_d^{*m}} - q^{*m}\|^2) \leq R_m d^{-2a_m} + \mathcal{E}_m \left(\frac{d}{n^m} + \frac{\|q\|^2}{n} \right).$$

Choosing a trade-off bandwidth $d_{opt} = Cn^{m/(2a_m+1)}$, we get a risk bound for $\mathbb{E}(\|\widehat{q_{d_{opt}}^{*m}} - q\|^2)$ of order $\max(n^{-2ma_m/(2a_m+1)}, n^{-1})$. If $2ma_m/(2a_m+1) \geq 1$, i.e. $2a_m(m-1) \geq 1$, the risk rate has order $1/n$. This occurs for instance if $m \geq 2$ and $a_m \geq 1/2$.

10.3 Estimation of the jump density

The Sobolev regularities of f and q with $q = q_\Delta$ are linked. Recall that for any function $h \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ we denote by h_d the function defined by $h_d^* = h^* 1_{[-\pi d, \pi d]}$.

Proposition 10.4 *Let the density f belong to $\mathcal{C}(a, L)$ (see (4.14)). Then q defined by (10.3) and (10.8) belongs to $\mathcal{C}(a, L)$. In particular,*

$$\|q\| \leq \|f\|.$$

Proof. Consider f integrable with $\|f\|_1 = \int |f|$ and square integrable such that $\int (1+x^2)^a |f^*(x)|^2 dx \leq L$. Then

$$\begin{aligned}
& \int (1+x^2)^a |q^*(x)|^2 dx \\
&= \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}} \right)^2 \sum_{m,k \geq 1} \frac{(c\Delta)^m}{m!} \frac{(c\Delta)^k}{k!} \int (1+x^2)^a [f^*(x)]^m [f^*(-x)]^k dx \\
&\leq \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}} \right)^2 \sum_{m,k \geq 1} \frac{(c\Delta)^m}{m!} \frac{(c\Delta)^k}{k!} \|f\|_1^{m+k-2} \int (1+x^2)^a |f^*(x)|^2 dx \\
&\leq L \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}} \right)^2 \frac{1}{\|f\|_1^2} \left(\sum_{m \geq 1} \frac{(c\Delta)^m}{m!} \|f\|_1^m \right)^2 \\
&= L \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}} \frac{\exp(c\Delta \|f\|_1) - 1}{\|f\|_1} \right)^2 := L(\Delta) < +\infty
\end{aligned}$$

As f is a density, $\|f\|_1 = 1$ and $L(\Delta) = L$. This implies the announced result for q .

We assume now that $c \in [c_0, c_1]$ with $c_1 \Delta \leq \log 2/2$ and consider the estimator $\widehat{f_{K,d}}$ given by

$$\widehat{f_{K,d}}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \widehat{c_m(\Delta)} \widehat{q_d^{*m}}(x). \quad (10.26)$$

where $\widehat{c_m(\Delta)}$ is the estimator of $c_m(\Delta)$ given in (10.11).

Proposition 10.5 *Assume that $c \in [c_0, c_1]$ with $c_0 > 0$ and $c_1 \Delta \leq \log 2/2$. Then the estimator $\widehat{f_{K,d}}$ is such that*

$$\mathbb{E}(\|\widehat{f_{K,d}} - f\|^2) \leq \frac{5}{2\pi} \int_{|t| \geq \pi d} |f^*(t)|^2 dt + \frac{10d}{n} + 5A_K \Delta^{2K+2} + \frac{5B_K}{n}, \quad (10.27)$$

with

$$A_K = 6 \frac{\|f\|^2}{(K+2)^2} (\sqrt{2}c)^{2K+2}, \quad (10.28)$$

$$B_K = 2(K+1)(1+2\|f\|^2) \left\{ C_1 + \Delta^2 \sum_{m=2}^{K+1} \frac{(C_m + 2^m c^{2(m-1)}) \mathcal{E}_m}{m^2} \Delta^{2(m-2)} \right\}, \quad (10.29)$$

where C_m, \mathcal{E}_m are the constants appearing respectively in (10.12) and in (10.17).

Proof. Recall that $f^* = \sum_{m \geq 1} ((-1)^{m+1}/m) c_m(\Delta) (q^*)^m$ (see (10.6)-(10.7)). Let f_d be such that $f_d^* = f^* 1_{[-\pi d, \pi d]}$ and $f_{K,d}$ be such that

$$f_{K,d}^* = 1_{[-\pi d, \pi d]} \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} c_m(\Delta) (q^*)^m.$$

Define

$$\widetilde{f_{K,d}}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} c_m(\Delta) \widehat{q_d^{*m}}(x), \quad \text{with } c_m(\Delta) = \frac{(e^{c\Delta} - 1)^m}{c\Delta}, \quad (10.30)$$

so that:

$$(\widehat{f_{K,d}})^* = 1_{[-\pi d, \pi d]} \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} c_m(\Delta) (\widehat{q^*})^m.$$

We distinguish the first term of this development from the other ones and set

$$\widehat{f_{K,d}} = \widehat{f_{K,d}}^{(1)} + \mathcal{R} \widehat{f_{K,d}}, \text{ with } \widehat{f_{K,d}}^{(1)} = c_1(\Delta) \widehat{q_d^{*1}} = c_1(\Delta) \widehat{q_d}. \quad (10.31)$$

Analogously, with q_d such that $q_d^* = q^* 1_{[-\pi d, \pi d]}$,

$$f_{K,d} = f_{K,d}^{(1)} + \mathcal{R} f_{K,d}, \text{ with } f_{K,d}^{(1)} = c_1(\Delta) q_d \quad (10.32)$$

The following decomposition of the \mathbb{L}^2 -norm holds:

$$\begin{aligned} \|f - \widehat{f_{K,d}}\| &\leq \|f - f_d\| + \|f_d - f_{K,d}\| + \|f_{K,d}^{(1)} - \widehat{f_{K,d}}^{(1)}\| \\ &\quad + \|\mathcal{R} f_{K,d} - \widehat{\mathcal{R} f_{K,d}}\| + \|\widehat{f_{K,d}} - \widehat{f_{K,d}}^{(1)}\|, \end{aligned}$$

which involves two bias terms and two stochastic error terms. The first bias term is the usual deconvolution bias term:

$$\|f - f_d\|^2 = \frac{1}{2\pi} \int_{|t| \geq \pi d} |f^*(t)|^2 dt$$

Noting that

$$f_d^* - f_{K,d}^* = 1_{[-\pi d, \pi d]} \sum_{m=K+2}^{\infty} \frac{(-1)^{m+1}}{m} c_m(\Delta) (q^*)^m,$$

we get, using that $|q^*(t)| \leq 1$ and $\|q\| \leq \|f\|$ (see Proposition 10.4):

$$\begin{aligned} 2\pi \|f_d - f_{K,d}\|^2 &= \|f_d^* - f_{K,d}^*\|^2 = \int_{-\pi d}^{\pi d} \left| \sum_{m=K+2}^{\infty} \frac{(-1)^{m+1}}{m} c_m(\Delta) (q^*)^m(t) \right|^2 dt \\ &\leq \int_{-\pi d}^{\pi d} \left(\sum_{m \geq K+2} \frac{1}{m} c_m(\Delta) |q^*(t)| \right)^2 dt \\ &\leq 2\pi \|q\|^2 \left(\sum_{m \geq K+2} \frac{1}{m} c_m(\Delta) \right)^2 \\ &\leq \frac{2\pi \|f\|^2}{(c\Delta)^2 (K+2)^2} \left(\frac{(e^{c\Delta} - 1)^{K+2}}{2 - e^{c\Delta}} \right)^2 \\ &\leq \frac{4\pi \|f\|^2 (\sqrt{2}c\Delta)^{2K+2}}{((K+2)^2 (2 - e^{2\Delta}))^2} \leq 2\pi A_K \Delta^{2K+2}, \quad (10.33) \end{aligned}$$

where in the last line, we have used $1/(2 - e^{c\Delta})^2 \leq 1/(2 - \sqrt{2})^2 \leq 3$ and $e^{c\Delta} - 1 \leq \sqrt{2}c\Delta$ and A_K is given in (10.28).

To study the next term, we recall that, $\mathbb{E}(|\widehat{(q^*)}(t) - (q^*)(t)|^2) \leq 1/n$. Then we get

$$\begin{aligned} 2\pi\mathbb{E}\left(\|f_{K,d}^{(1)} - \widetilde{f_{K,d}^{(1)}}\|^2\right) &= \int_{-\pi d}^{\pi d} \mathbb{E}\left(\left|c_1(\Delta)[\widehat{(q^*)}(t) - (q^*)(t)]\right|^2\right) dt \\ &\leq \frac{2\pi d [c_1(\Delta)]^2}{n} \leq \frac{4\pi d}{n} \end{aligned} \quad (10.34)$$

since $c_1(\Delta) \leq \sqrt{2}$.

Hereafter, we use inequality (10.17) of Proposition 10.3.

$$\begin{aligned} &2\pi\mathbb{E}\left(\|\mathcal{R}f_{K,d} - \widetilde{\mathcal{R}f_{K,d}}\|^2\right) \\ &= \int_{-\pi d}^{\pi d} \mathbb{E}\left(\left|\sum_{m=2}^{K+1} \frac{(-1)^{m+1}}{m} c_m(\Delta)[\widehat{(q^*)^m}(t) - (q^*)^m(t)]\right|^2\right) dt \\ &\leq \int_{-\pi d}^{\pi d} (K+1) \sum_{m=2}^{K+1} \frac{1}{m^2} [c_m(\Delta)]^2 \mathbb{E}\left(|\widehat{(q^*)^m}(t) - (q^*)^m(t)|^2\right) dt \\ &\leq 2\pi K \sum_{m=2}^{K+1} \frac{\mathcal{E}_m}{m^2} [c_m(\Delta)]^2 \left(\frac{d}{n^m} + \frac{\|q\|^2}{n}\right) \end{aligned}$$

This yields, since $c_m(\Delta) \leq (\sqrt{2})^m (c\Delta)^{m-1}$ and $d/n \leq 1$,

$$\mathbb{E}\left(\|\mathcal{R}f_{K,d} - \widetilde{\mathcal{R}f_{K,d}}\|^2\right) \leq \frac{D_K}{n} \quad (10.35)$$

with

$$D_K = K \sum_{m=2}^{K+1} \frac{2^m c^{2(m-1)} \mathcal{E}_m}{m^2} \Delta^{2(m-1)} \left(\frac{1}{n^{m-2}} + \|q\|^2\right)$$

For the last term, we use Proposition 10.2, with the fact that the estimators $\widehat{c_m(\Delta)}$ and $\widehat{(q^*)^m}(t)$ are independent, and write

$$\begin{aligned}
& 2\pi\mathbb{E}\left(\|\widehat{f_{K,d}} - \widetilde{f_{K,d}}\|^2\right) \\
&= \int_{-\pi d}^{\pi d} \mathbb{E}\left(\left|\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} (\widehat{c_m(\Delta)} - c_m(\Delta)) (\widehat{q^*}^m(t)\right|^2 dt\right) \\
&\leq 2 \int_{-\pi d}^{\pi d} \mathbb{E}\left(\left|\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} (\widehat{c_m(\Delta)} - c_m(\Delta)) [(\widehat{q^*}^m(t) - (q^*)^m(t))\right|^2 dt\right) \\
&\quad + 2 \int_{-\pi d}^{\pi d} \mathbb{E}\left(\left|\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} (\widehat{c_m(\Delta)} - c_m(\Delta)) (q^*)^m(t)\right|^2 dt\right) \\
&\leq 2(K+1) \sum_{m=1}^{K+1} \frac{1}{m^2} \left\{ \mathbb{E}\left[(\widehat{c_m(\Delta)} - c_m(\Delta))^2\right] \int_{-\pi d}^{\pi d} \mathbb{E}\left[|(\widehat{q^*}^m(t) - (q^*)^m(t))|^2\right] dt \right. \\
&\quad \left. + \mathbb{E}\left[(\widehat{c_m(\Delta)} - c_m(\Delta))^2\right] \int_{-\pi d}^{\pi d} |q^*(t)|^{2m} dt \right\} \\
&\leq 2(K+1) \left\{ \frac{C_1}{n} \left(\frac{2\pi d}{n} + 2\pi\|q\|^2\right) \right. \\
&\quad \left. + \sum_{m=2}^{K+1} \frac{C_m \Delta^{2(m-1)}}{m^2} \left[\frac{\mathcal{E}_m}{n} \int_{-\pi d}^{\pi d} \left(\frac{1}{n^m} + \frac{1}{n} |q^*(t)|^2\right) dt + \frac{1}{n} \|q^*\|^2 \right] \right\}.
\end{aligned}$$

Therefore

$$2\pi\mathbb{E}\left(\|\widehat{f_{K,d}} - \widetilde{f_{K,d}}\|^2\right) \leq \frac{2\pi E_K}{n} \quad (10.36)$$

using that $d/n \leq 1$ and

$$E_K = 2(K+1) \left[C_1(1 + \|q\|^2) + \sum_{m=2}^{K+1} \frac{C_m}{m^2} \Delta^{2(m-1)} \mathcal{E}_m \left(\frac{1}{n^{m-1}} + 2\|q\|^2\right) \right].$$

This ends the proof of the result with $D_K + E_K \leq B_K$ and $\|q\| \leq \|f\|$.

If $f \in \mathcal{C}(a, L)$, choosing $d = d^* \propto n^{-1/(2a+1)}$, inequality (10.27) yields

$$\mathbb{E}(\|\widehat{f_{K,d^*}} - f\|^2) \leq Cn^{-2a/(2a+1)} + 5A_K \Delta^{2K+2}. \quad (10.37)$$

Usually, in high frequency data for continuous time models, rates are measured in terms of the total length time of observation which is here equal to $S_n \Delta$. Evaluating this random value as n tends to infinity, Δ tends to 0, we get that

$$S_n \Delta = \frac{S_n}{n} n \Delta \sim \frac{\Delta}{p(\Delta)} n \sim \frac{n}{c}.$$

The total length time of observation is asymptotically equivalent to n . For $n \Delta^{2K+2} \leq 1$, the result is comparable to the one obtained in Proposition 4.4 with a weaker constraint on Δ which now depends on K .

As in Section 4, we propose an adaptive selection procedure for choosing the cut-off parameter d in a restricted set $\{1, \dots, L_n\}$ with $L_n \leq n$. Let

$$\hat{d} = \arg \min_{1 \leq d \leq L_n} \{-\|\widehat{f_{K,d}}\|^2 + \text{pen}(d)\}, \text{ with } \text{pen}(d) = \kappa \frac{d}{n}.$$

We can prove the following result.

Theorem 10.1 *Assume that f is bounded and $L_n \leq n$. There exists a numerical value κ_0 such that for any κ larger than κ_0 , we get,*

$$\begin{aligned} \mathbb{E}(\|\widehat{f_{K,\hat{d}}} - f\|^2) &\leq 4 \min_{1 \leq d \leq L_n} \{\|f - f_d\|^2 + \text{pen}(d)\} \\ &\quad + 32A_K \Delta^{2K+2} + 32 \frac{B_K}{n} + \frac{C'}{n}, \end{aligned} \quad (10.38)$$

where C' is a constant.

Comparing the above inequality with (10.27), we see that the estimator is adaptive as its risk automatically realizes the best compromise between the squared bias term (first one, inside the min) and the variance term (second one, inside the min). The last two terms are standardly negligible. For the term $32A_K \Delta^{2K+2}$, either the sampling interval Δ for given K is tuned to make it negligible ($O(1/n)$) or n , Δ are given and K is chosen so that $n\Delta^{2K+2} \simeq 1$.

Using the estimator \hat{c} given in (10.11), we can conclude for the Lévy density.

Corollary 10.1 *Let $n(x) = cf(x)$ and $\hat{n}_{K,d}(x) = \hat{c}\widehat{f_{K,d}}(x)$ with \hat{c} given in (10.11). Then under the Assumptions of Theorem 10.1,*

$$\mathbb{E}(\|\hat{n}_{K,\hat{d}} - n\|^2) \leq 3c^2 \mathbb{E}(\|\widehat{f_{K,\hat{d}}} - f\|^2) + \frac{C''}{n}.$$

The corollary is straightforwardly obtained by writing

$$\hat{n}_{K,d} - n = c(\widehat{f_{K,d}} - f) + (\hat{c} - c)f + (\hat{c} - c)(\widehat{f_{K,d}} - f).$$

Then the bound follows from Proposition 10.2 and Theorem 10.1.

Proof of Theorem 10.1. We use the subspaces of S_d introduced in (4.15) to show that the estimators $\widehat{f_{K,d}}, l \leq L_n$ are minimizers of a projection contrast. The difference here from definition (4.17) is that we need the maximal space S_{L_n} in the contrast definition. Let

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{f_{K,L_n}} \rangle.$$

Note that, for $d \leq L_n$ and $t \in S_d$, $\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{f_{K,d}} \rangle$, and

$$\arg \min_{t \in S_d} \gamma_n(t) = \widehat{f_{K,d}}, \quad \text{with} \quad \gamma_n(\widehat{f_{K,d}}) = -\|\widehat{f_{K,d}}\|^2.$$

Now, the steps of Theorem 4.1 can be followed. For $d, d^* \leq L_n$, $s \in S_d$ and $t \in S_{d^*}$:

$$\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - 2\langle t - s, \widehat{f_{K,L_n}} - f \rangle$$

and $\langle t - s, \widehat{f_{K,L_n}} - f \rangle = \langle t - s, \widehat{f_{K,L_n}} - f_{L_n} \rangle$. By definition of \hat{d} ,

$$\gamma_n(\widehat{f_{K,\hat{d}}}) + \text{pen}(\hat{d}) \leq \gamma_n(\widehat{f_{K,d}}) + \text{pen}(d) \leq \gamma_n(f_d) + \text{pen}(d).$$

Thus, we obtain, $\forall d \in \{1, \dots, L_n\}$,

$$\begin{aligned} \|\widehat{f_{K,\hat{d}}} - f\|^2 &\leq \|f_d - f\|^2 + \text{pen}(d) + 2\langle \widehat{f_{K,\hat{d}}} - f_d, \widehat{f_{K,L_n}} - f_{L_n} \rangle - \text{pen}(\hat{d}) \\ &\leq \|f_d - f\|^2 + \text{pen}(d) + \frac{1}{4}\|\widehat{f_{K,\hat{d}}} - f_d\|^2 \\ &\quad + 4 \sup_{t \in S_d + S_{\hat{d}}, \|t\|=1} \langle t, \widehat{f_{K,L_n}} - f_{L_n} \rangle^2 - \text{pen}(\hat{d}) \end{aligned} \quad (10.39)$$

Then

$$\frac{1}{4}\|\widehat{f_{K,\hat{d}}} - f_d\|^2 \leq \frac{1}{2}\|\widehat{f_{K,\hat{d}}} - f\|^2 + \frac{1}{2}\|f - f_d\|^2. \quad (10.40)$$

Now, we use the specific decompositions (10.31) and (10.32):

$$\begin{aligned} \langle t, \widehat{f_{K,L_n}} - f_{L_n} \rangle &= \langle t, \widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}} \rangle + \langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)} \rangle \\ &\quad + \langle t, \mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n} \rangle + \langle t, f_{K,L_n} - f_{L_n} \rangle. \end{aligned}$$

By the Cauchy-Schwarz Inequality and for $\|t\| = 1$, we have

$$\begin{aligned} \langle t, \widehat{f_{K,L_n}} - f_{L_n} \rangle^2 &\leq 4\|\widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}}\|^2 + 4\|\mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n}\|^2 \\ &\quad + 4\|f_{K,L_n} - f_{L_n}\|^2 + 4\langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)} \rangle^2. \end{aligned} \quad (10.41)$$

Thus, inserting (10.40) and (10.41) in (10.39) yields

$$\begin{aligned} \frac{1}{2}\|\widehat{f_{K,\hat{d}}} - f\|^2 &\leq \frac{3}{2}\|f_d - f\|^2 + 16\|f_{K,L_n} - f_{L_n}\|^2 \\ &\quad + 16\|\widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}}\|^2 + 16\|\mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n}\|^2 + \text{pen}(d) \\ &\quad + 16 \sup_{t \in S_{d \vee \hat{d}}, \|t\|=1} \langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)} \rangle^2 - \text{pen}(\hat{d}) \end{aligned}$$

Here, the bounds of Proposition 10.5 can be applied. Indeed (10.33), (10.35) and (10.36) are uniform with respect to d and imply

$$\|f_{K,L_n} - f_{L_n}\|^2 \leq A_K \Delta^{2(K+2)}, \quad \mathbb{E}(\|\mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n}\|^2) \leq D_K/n,$$

and

$$\mathbb{E}(\|\widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}}\|^2) \leq E_K/n.$$

Below, we prove the following inequality which is to be compared with Lemma 4.1:

$$\mathbb{E} \left(\sup_{t \in S_{d \vee \hat{d}}, \|t\|=1} \langle t, \widetilde{f_{K, L_n}}^{(1)} - f_{K, L_n}^{(1)} \rangle^2 - p(d, \hat{d}) \right)_+ \leq \frac{C'}{n}, \quad (10.42)$$

where $p(d, d') = 8d \vee d' / n$ and $16p(d, d') \leq \text{pen}(d) + \text{pen}(d')$ as soon as $\kappa \geq \kappa_0 = 16 \times 8$.

Consequently,

$$\mathbb{E}(16p(d, \hat{d}) - \text{pen}(\hat{d})) \leq \text{pen}(d)$$

and

$$\mathbb{E}(\|\widehat{f_{K, \hat{d}}} - f\|^2) \leq 4\|f - f_d\|^2 + 4\text{pen}(d) + 32A_K \Delta^{2(K+2)} + 32\frac{B_K}{n} + \frac{32C'}{n}.$$

Proof of (10.42). We consider $t \in S_{d^*}$ for $d^* = d \vee d'$ with $d, d' \leq L_n$ and (see (10.31) and (10.32))

$$v_n(t) = \langle t, \widetilde{f_{K, L_n}}^{(1)} - f_{K, L_n}^{(1)} \rangle = c_1(\Delta) \langle t, \hat{q}_{L_n} - q_{L_n} \rangle = \frac{1}{n} \sum_{k=1}^n (\psi_t(Z_k) - \mathbb{E}(\psi_t(Z_k)))$$

where

$$\psi_t(z) = \frac{c_1(\Delta)}{2\pi} \int t^*(u) e^{iuz} du = c_1(\Delta) t(z).$$

We apply the Talagrand Inequality (see Appendix). To this aim, we compute the quantities M, H, v . First

$$\sup_{t \in S_{d^*}, \|t\|=1} \sup_z |\psi_t(z)| \leq \frac{c_1(\Delta)}{2\pi} \sqrt{2\pi d^*} \times \sup_{t \in S_{d^*}, \|t\|=1} \|t^*\| = c_1(\Delta) \sqrt{d^*} := M.$$

The density of Z_1 is q which satisfies

$$\|q\|_\infty \leq \sum_{m \geq 1} \frac{1}{e^{c\Delta} - 1} \frac{(c\Delta)^m}{m!} \|f^{*m}\|_\infty \leq \|f\|_\infty.$$

Therefore,

$$\sup_{t \in S_{d^*}, \|t\|=1} \text{Var}(\psi_t(Z_1)) \leq c_1^2(\Delta) \times \sup_{t \in S_{d^*}, \|t\|=1} \mathbb{E}(t^2(Z_1)) \leq c_1^2(\Delta) \|f\|_\infty := v.$$

Lastly, using the bound in (10.34) and the fact that for $t \in S_{d^*}$,

$$\langle t, \widetilde{f_{K, L_n}}^{(1)} - f_{K, L_n}^{(1)} \rangle = \langle t, \widetilde{f_{K, d^*}}^{(1)} - f_{K, d^*}^{(1)} \rangle,$$

we get

$$\begin{aligned}\mathbb{E}\left(\sup_{t \in S_{d^*}, \|t\|=1} v_n^2(t)\right) &= \mathbb{E}\left(\sup_{t \in S_{d^*}, \|t\|=1} \langle t, \widetilde{f_{K,d^*}}^{(1)} - f_{K,d^*}^{(1)} \rangle^2\right) \\ &\leq \mathbb{E}\left(\|\widetilde{f_{K,d^*}}^{(1)} - f_{K,d^*}^{(1)}\|^2\right) \leq \frac{2d^*}{n} := H^2.\end{aligned}$$

Therefore, Lemma .1 yields with $\varepsilon^2 = 1/2$,

$$\mathbb{E}\left(\sup_{t \in S_{d^*}, \|t\|=1} v_n^2(t) - 4H^2\right) \leq \frac{A_1}{n} (e^{-A_2 d^*} + e^{-A_3 \sqrt{n}})$$

for constants A_1, A_2, A_3 depending on $c_1(\Delta)$ and $\|f\|_\infty$. Now since

$$\sum_{d'=1}^{L_n} e^{-A_2 d \vee d'} = d e^{-A_2 d} + \sum_{d < d' \leq L_n} e^{-A_2 d'}$$

is bounded by say B_2 and $L_n e^{-A_3 \sqrt{n}}$ is bounded by B_3 , we get

$$\mathbb{E}\left(\sup_{t \in S_{d \vee \hat{d}}, \|t\|=1} v_n^2(t) - 8 \frac{d \vee \hat{d}}{n}\right) \leq \sum_{d'} \mathbb{E}\left(\sup_{t \in S_{d \vee d'}, \|t\|=1} v_n^2(t) - 4H^2\right) \leq \frac{B_4}{n}.$$

This ends the proof of (10.42) and thus of Theorem 10.1. \square

10.4 Simulations

We have implemented the adaptive estimator on different examples of jump densities f , namely,

1. A Gaussian $\mathcal{N}(0, 1)$.
2. A mixture of a Gaussian and a Gamma $\frac{2}{3}\mathcal{N}(-4, 1) + \frac{1}{3}\Gamma(3, 1)$.
3. A Laplace $L(0, 1)$ with density $\exp(-|x|)/2$.
4. A Gamma $\Gamma(5, 1)$.

After preliminary experiments the constant κ is taken equal to 17.6 and the cut-off \hat{d} is selected among 100 equispaced values between 0 and 10. We consider different values of Δ : 0.2, 0.5, 0.8. For each Δ we choose K such that $n\Delta^{2K+2} \leq 1$; more precisely the corresponding values of K are 2, 5, 17 respectively.

Results are given in Figure 6, where 50 estimated curves are plotted on the same figure to show the small variability of the estimator. We take a sample size $n = 5000$ and an intensity $c = 0.5$, the first lines give the result for $\Delta = 0.2$ ($K = 2$), the second for $\Delta = 0.5$ ($K = 5$) and the last for $\Delta = 0.8$ ($K = 17$). On top of each graph we give the mean of selected values for \hat{d} and the associated standard deviation in parenthesis evaluated over the fifty plots given. It appears that for each Δ the

estimator reproduces well the estimated density with little variability. Increasing Δ , and therefore K , does not affect the accuracy nor the variability of the estimator.

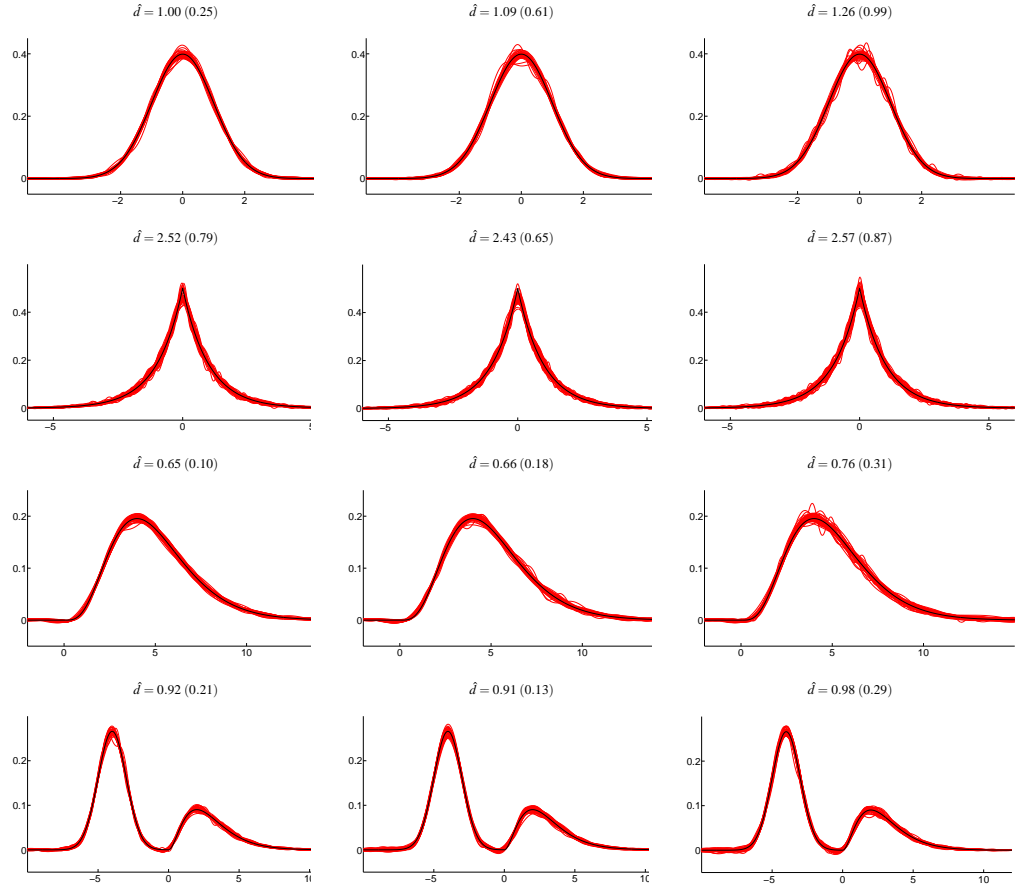


Fig. 6 Estimation of the jump density f for a Gaussian $\mathcal{N}(0, 1)$ (first line), Laplace $L(0, 1)$ second line, Gamma $\Gamma(5, 1)$ (third line) and the mixture $\frac{2}{3}\mathcal{N}(-4, 1) + \frac{1}{3}\Gamma(3, 1)$ (fourth line) with $c = 0.5$ and $n = 5000$. True density (bold black line) and 50 estimated curves (red lines), left $\Delta = 0.2$ and $K = 2$; middle $\Delta = 0.5$ and $K = 5$; right $\Delta = 0.8$ and $K = 17$. The value \hat{d} is the mean over the 50 selected \hat{d} 's (with standard deviation in parenthesis).

11 Bibliographic comments

We give here some bibliographic comments which are far from exhaustive and focus mainly on our text.

Adaptive nonparametric methods have been developed for density estimation from *i.i.d.* observations: see [24] for wavelet thresholding methods, [6] or [53] for model selection and contrast penalization methods or [34] for adaptive bandwidth selection in kernel estimation. In the present chapter, we have adapted some of these approaches for estimating the Lévy density.

For *i.i.d.* data contaminated with additive noise, specific methods have been introduced, based on Fourier inversion and called deconvolution methods. In the first papers, the noise distribution is assumed to be known, see [29] for nonadaptive kernel, [57] for adaptive wavelet estimator and [22] for adaptive cut-off selection. More recently, the case of unknown noise distribution has been considered, see [55], [41], [21], [44]. The estimation of the Lévy density for Lévy processes relies on the explicit form of the characteristic function and thus takes inspiration in the deconvolution methods.

Lévy processes have been increasingly used for modelling financial data (see *e.g.* [11], [52], [26], [2], [8] and [3], [15]). The nonparametric estimation of the Lévy density has been studied for a continuous time observation of the sample path on a time interval $[0, T]$ with T tending to infinity ([33]) or for discrete time observations. In the latter case, authors distinguish between low frequency data (sampling interval Δ is fixed) or high frequency data (Δ tends to 0). We concentrate in this chapter on high frequency data setting since it is simpler and allows to consider several adaptive estimation methods: deconvolution with cut-off selection, contrast penalization, see our works [17], [19], [20], and also [30], [31], [62] and adaptive kernels (see Section 4.3, and also [7]).

The nonparametric estimation in the case of low frequency observations is more difficult and closely related to a deconvolution problem with estimated noise density, see [56], [35], [13], [18], [36], [45].

Section 10 is specific to compound Poisson processes, widely used in insurance modelling, see [27], and to the problem of decompounding (see [12]). The discretized observation is defined as in [28], to take into account that null increments do not bring information on the jump density. The present approach is an improvement of [25].

The chapter only deals with upper risk bounds, but to check the optimality of the estimators, lower bounds are needed: they are provided, in the high frequency setting by [31], [7], and in the low frequency setting by [8], [56], [46], [45]. Lower bound in the specific case of decompounding is obtained in [25].

Acknowledgements If you want to include acknowledgments of assistance please do it here.

The Talagrand inequality. The result below follows from the Talagrand concentration inequality given in [49] and arguments in [10] (see the proof of their Corollary 2 page 354).

Lemma .1 (*Talagrand Inequality*) Let Y_1, \dots, Y_n be independent random variables, let $v_{n,Y}(f) = (1/n) \sum_{i=1}^n [f(Y_i) - \mathbb{E}(f(Y_i))]$ and let \mathcal{F} be a countable class of uniformly bounded measurable functions. Then for $\varepsilon^2 > 0$

$$\mathbb{E} \left[\sup_{f \in \mathcal{F}} |v_{n,Y}(f)|^2 - 2(1 + 2\varepsilon^2)H^2 \right]_+ \leq \frac{4}{K_1} \left(\frac{v}{n} e^{-K_1 \varepsilon^2 \frac{nH^2}{v}} + \frac{98M^2}{K_1 n^2 C^2(\varepsilon^2)} e^{-\frac{2K_1 C(\varepsilon^2) \varepsilon}{7\sqrt{2}} \frac{nH}{M}} \right),$$

with $C(\varepsilon^2) = \sqrt{1 + \varepsilon^2} - 1$, $K_1 = 1/6$, and

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M, \quad \mathbb{E} \left[\sup_{f \in \mathcal{F}} |v_{n,Y}(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{k=1}^n \text{Var}(f(Y_k)) \leq v.$$

By standard density arguments, this result can be extended to the case where \mathcal{F} is a unit ball of a linear normed space, after checking that $f \mapsto v_n(f)$ is continuous and \mathcal{F} contains a countable dense family.

The Rosenthal inequality. (see e.g. [37]) Let $(X_i)_{1 \leq i \leq n}$ be n independent centered random variables, such that $\mathbb{E}(|X_i|^p) < +\infty$ for an integer $p \geq 1$. Then there exists a constant $C(p)$ such that

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

The Young inequality. (see [39]) Let f be a function belonging to $\mathbb{L}^p(\mathbb{R})$ and g belonging to $\mathbb{L}^q(\mathbb{R})$, let p, q, r be real numbers in $[1, +\infty]$ and such that

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1.$$

Then

$$\|f * g\|_r \leq \|f\|_p \|g\|_q.$$

where $f * g$ is the convolution product and $\|f\|_p^p = \int |f(x)|^p dx$. In particular, for $p = 1, r = q = 2$, we have $\|f * g\|_2 \leq \|f\|_1 \|g\|_2$.

References

1. Aït-Sahalia Y. and Jacod J. (2007). Volatility estimators for discretely sampled Lévy processes. *The Annals of Statistics* **35**, 355-392.
2. Barndorff-Nielsen O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.* **24**, 1-13.
3. Barndorff-Nielsen O.E. and Shephard N. (2001). Modelling by Lévy processes for financial econometrics. In: Lévy processes. Theory and Applications (Barndorff-Nielsen O.E., Mikosch T., Resnick S.L.), 283-318.

4. Barndorff-Nielsen, O. E. and Shephard, N. (2006). Multipower variation and stochastic volatility. *Stochastic finance*, 73-82, Springer, New York.
5. Barndorff-Nielsen O.E., Shephard N. and Winkel M. (2006). Limit theorems for multipower variation in the presence of jumps. *Stochastic Process. Appl* **116**, 796-806.
6. Barron, A.R., Birgé, L. and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields.* **97**, 113-150.
7. Bec, M., Lacour, C. (2012) Adaptive kernel estimation of the Lévy density, Preprint arXiv:1205.4692.
8. Belometsny D. and Reiss M. (2006). Spectral calibration of exponential Lévy models. *Finance and Stochastics* **10** 449-474.
9. Bertoin J. (1996). *Lévy processes*. Cambridge University Press.
10. Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
11. Blattberg R. and Gonedes N. (1974). A comparison of the stable and Student distributions as statistical models for stock prices. *J. Business* **47**, 244-280.
12. Buchmann, B. and Grübel, R. (2003) Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.* **31** 1054-1074.
13. Chen, S. X., Delaigle, A. and Hall, P. (2010) Nonparametric estimation for a class of Lévy processes. *J. Econometrics* **157**, 257-271.
14. Chesneau, C., Comte, F. and Navarro, F. (2013) Fast nonparametric estimation for convolutions of densities. *Canadian J. Statist.* **41**, 617-636.
15. Cont R. and Tankov P. (2004). Financial modelling with jump processes. In: Financial Mathematics Series, Chapman & Hall, CRC, Boca Raton.
16. Comte, F., Duval, C. and Genon-Catalot, V. (2013). Nonparametric density estimation in compound Poisson processes using convolution power estimators. To appear in *Metrika*.
17. Comte, F. and Genon-Catalot (2009) Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications* **119**, 4088-4123.
18. Comte, F. and Genon-Catalot, V. (2010) Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **46**, 595-617.
19. Comte, F. and Genon-Catalot, V. (2010) Non-parametric estimation for pure jump irregularly sampled or noisy Lévy processes. *Stat. Neerl.* **64**, 290-313.
20. Comte, F. and Genon-Catalot, V. (2011) Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.* **39**, 803-837.
21. Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73**, 601-627.
22. Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006) Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34**, 431-452.
23. DeVore, R.A. and Lorentz, G.G. (1993). *Constructive approximation*. Springer-Verlag.
24. Donoho, D., Johnstone, I., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.
25. Duval, C. (2013) Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.* **123**, 3963-3986.
26. Eberlein E. and Keller U. (1995). Hyperbolic distributions in finance. *Bernoulli*, **1**, 281-299.
27. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling extremal events. For insurance and finance. *Applications of Mathematics* **33**. Springer-Verlag, Berlin, 1997.
28. van Es, B., Gugushvili, S. and Spreij, P. (2007) A kernel type nonparametric density estimator for decompounding. *Bernoulli* **13**, 672-694.
29. Fan J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, no. **3**, 1257-1272.
30. Figueroa-López J.E. (2008). Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.* **78**, 3355-3365.
31. Figueroa-López J.E. (2009). Nonparametric estimation for Lévy models based on discrete-sampling. *IMS Lecture Notes-Monograph Series. Optimality: The Third Erich L. Lehmann Symposium*, **57**, 117-146.

32. Figueroa-López, J. E. (2011) Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli* **17**, 643-670.
33. Figueroa-López, J. E. and Houdré, C. (2006) Risk bounds for the non-parametric estimation of Lévy processes. *High dimensional probability*, 96-116, IMS Lecture Notes Monogr. Ser., 51, Inst. Math. Statist., Beachwood, OH.
34. Goldenshluger A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, no. 3, 1608-1632.
35. Gugushvili, S. (2009) Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21**, 321-343.
36. Gugushvili, S. (2012) Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, 282-307.
37. Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its applications*. Academic press, Inc. (London) LTD.
38. Härdle, W., Kerkycharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. Lecture Notes in Statistics, 129. Springer-Verlag, New York.
39. Hirsch, F. and Lacombe, G. (1999). *Elements of functional analysis*. Graduate Texts in Mathematics, 192. Springer-Verlag, New York.
40. Jacod J. (2007). Asymptotic properties of power variations of Lévy processes. *ESAIM: Probability and Statistics*, **11**, 173-196.
41. Johannes, J. (2009) Deconvolution with unknown error distribution. *Ann. Statist.* **37**, 2301-2323.
42. Jongbloed, G. and van der Meulen, F. H. (2006) Parametric estimation for subordinators and induced OU processes. *Scand. J. Statist.* **33**, 825-847.
43. Jongbloed, G., van der Meulen, F. H. and van der Vaart, A. W. (2005) Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **11**, 759-791.
44. Kappus, J. (2012) J. Kappus, Nonparametric adaptive estimation for discretely observed Lévy processes, Ph.D. Thesis, Humboldt Universität zu Berlin.
45. Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.*, **124**, 730-758.
46. Kappus, J. and Reiss, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.* **64**, 314-328.
47. Katz, R. W. (2002) Stochastic modeling of hurricane damage. *J. Appl. Meteorol.* **41**, 754-762.
48. Kerkycharian, G., Lepski, O. and Picard, D. (2001) Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*. 121 (2001), **2**, 137-170.
49. Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.
50. Küchler, U. and Tappe, S. (2008) Bilateral Gamma distributions and processes in financial mathematics. *Stochastic Processes and Their Applications*, **118**, 261-283.
51. Lebedev, N.N. (1972). *Special functions and their applications*. Revised edition, translated from the Russian and edited by Richard A. Silverman. Dover Publications, Inc., New York.
52. Madan, D.B. and Seneta, E. (1990) The Variance Gamma (V.G.) Model for Share Market Returns. *The Journal of Business* **63**, No. 4, 511-524.
53. Massart, P. (2007) *Concentration inequalities and model selection*. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003. With a foreword by Jean Picard. Lecture Notes in Mathematics, 1896. Springer, Berlin
54. Meyer, Y. (1990). *Ondelettes et opérateurs*. I. Hermann, Paris.
55. Neumann, M. (1997) On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7**, 307-330.
56. Neumann, M. H. and Reiss, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15**, 223-248.
57. Pensky, M. and Vidakovic, B (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033-2053.
58. Sato K.I. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge Studies in Advanced Mathematics, 68. Cambridge University Press, Cambridge.

59. Scalas, E. (2006). The application of continuous-time random walks in finance and economics. *Physica A* **362**, 225-239.
60. Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, **16**, 925-935.
61. Tsybakov, A. (2009) Introduction to Nonparametric Estimation. Springer.
62. Ueltzhöfer, F. A. J. and Klüppelberg, C. (2011). An oracle inequality for penalised projection estimation of Lévy densities from high-frequency observations. *J. Nonparametr. Stat.* **23**, 967-989.
63. Woerner J.H.C. (2006). Power and Multipower variation: inference for high frequency data. *Stochastic Finance*, Springer, New York, 343364.