Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 123 (2013) 3963–3986

stochastic processes and their applications

www.elsevier.com/locate/spa

Density estimation for compound Poisson processes from discrete data

Céline Duval*

CREST and Université Paris-Dauphine, CNRS-UMR 7534, France

Received 25 March 2013; received in revised form 13 June 2013; accepted 13 June 2013
Available online 27 June 2013

Abstract

In this article we investigate the nonparametric estimation of the jump density of a compound Poisson process from the discrete observation of one trajectory over $[0, T]$. We consider the case where the sampling rate $\Delta = \Delta_T \to 0$ as $T \to \infty$. We propose an adaptive wavelet threshold density estimator and study its performance for L_p losses, $p \geq 1$, over Besov spaces. The main novelty is that we achieve minimax rates of convergence for sampling rates Δ_T that vanish slowly. The estimation procedure is based on the explicit inversion of the operator giving the law of the increments as a nonlinear transformation of the jump density.

MSC: 62G99; 62M99; 60G50

Keywords: Compound Poisson process; Discretely observed random process; Decompounding; Wavelet density estimation

1. Introduction

1.1. Setting and motivation

Let R be a standard homogeneous Poisson process with intensity ϑ in $(0, \infty)$; we define the compound Poisson process X as

$$X_t = \sum_{i=1}^{R_t} \xi_i, \quad t \geq 0$$

* Tel.: +33 141173866.
E-mail address: duval@ceremade.dauphine.fr.

0304-4149/$ - see front matter
http://dx.doi.org/10.1016/j.spa.2013.06.006

where (ξ_i) are independent and identically distributed random variables with density f and independent of the Poisson process R. We discretely observe the process X over $[0, T]$ at times $i\Delta$, for some $\Delta > 0$ and $i \in \mathbb{N}$,

$$\left(X_{\Delta}, \ldots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\right). \tag{1}$$

In this paper, we let $\Delta = \Delta_T \to 0$ as $T \to \infty$. The assumption $T \to \infty$ is necessary to observe infinitely many jumps; the asymptotic $\Delta_T \to 0$ is often referred to as high frequency data. In this statistical setting, the estimation problem for a discretely observed compound Poisson process has been widely studied. A compound Poisson process is a particular pure jump Lévy process and can be studied accordingly using the Lévy–Khintchine formula as in Figueroa-López [19], Comte and Genon-Catalot [10,12] or Bec and Lacour [4]. Their estimators are proven minimax for L_2 loss functions provided the sampling rate tends rapidly to 0 (see the discussion in Section 4). However, a methodology based on the Lévy–Khintchine formula requires to work with L_2 losses and only works for sampling rates such that $T\Delta^2 \leq 1$. For compound Poisson processes another possible approach is decompounding. It consists in inverting the operator giving the law of the increments as a nonlinear transformation of the jump density.

Decompounding methods have been introduced by Buchmann and Grübel [6] to estimate a compound Poisson process observed at a fixed sampling rate ($\Delta = 1$). They focus on discrete compound laws but also provide a method to estimate the distribution function of the jumps in the continuous case (see Section 4). When the intensity ϑ is known, their estimator converges to a Gaussian process with known covariance structure. However they provide no uniform bound for the estimation error. Optimality is not investigated.

In this paper we apply a decompounding method to estimate the jump density f, when the intensity ϑ is unknown. This approach gives new insights on this problem.

1. It removes the constraint $T\Delta^2 \leq 1$: our estimator is minimax for any Δ polynomially decreasing to 0 with T (*i.e.* Δ is of the order of $T^{-\delta}$ for some $\delta > 0$) and is consistent if Δ decays even slower.
2. It allows the use of wavelet type density estimators; we study the rate of convergence for L_p loss functions, $p \geq 1$, and over Besov classes of densities.

Compound Poisson processes are widely used to model phenomena where random events occur at random discrete times: in biology (see *e.g.* Huelsenbeck et al. [23] or Boys et al.), in statistical physics to model earthquakes (see *e.g.* Moharir [31]) or rainfall (see *e.g.* Alexandersson [1]), in ecology for species counts (see *e.g.* Etienne et al. [17]), and by insurance companies to model big claims of subscribers (see *e.g.* the Cramér–Lundberg model in Embrechts et al. [16], Katz [24], Scalas [34] or Mikosch [30]). They are also encountered in finance to model either asset prices (see *e.g.* Repetowicz et al. [33] or Masoliver et al. [28]) or the order book (see Avellaneda and Stoikov [3], Avellaneda et al. [2], Cont and de Larrard [13] or Guilbaud and Pham [20]).

In practice the choice of the sampling rate Δ is central and affects optimal procedures: unexpected phenomena, like efficiency loss, can be observed when the sampling rate varies (see Duval and Hoffmann [15]). In most applied fields exhaustive samples (collection of the jumps and their arrival times) are not available (see *e.g.* Boys et al. [5], rainfall data are collected daily (Alexandersson [1])). Even in finance, where exhaustive samples are usually available, subsampling schemes are encountered, for instance due to computer limitations. The following question may arise: *if the number of observations $n = \lfloor T\Delta^{-1}\rfloor$ is imposed, how should Δ be chosen?* With a ratio T/Δ fixed, increasing Δ enables to observe the process on a longer time interval T. Since the error of optimal estimators decreases with T (see Theorem 2.1 hereafter or Comte and Genon-Catalot [10] for instance), increasing Δ enables to increase T and then reduces the estimation error. This motivates the construction of an estimation procedure, which is optimal for sampling rates going slowly to 0.

1.2. Our results

Notation and assumption. We investigate the nonparametric estimation of the density f, from the observations (1), on a compact interval $\mathcal{D}$ included in $\mathbb{R}$. We measure the estimation accuracy, uniformly over Besov balls, by the mean of the following L_p loss function

$$\left(\mathbb{E}\left[\|\widehat{f}-f\|_{L_p(\mathcal{D})}^p\right]\right)^{1/p}, \tag{2}$$

where $\widehat{f}$ is an estimator of f, $p \geq 1$ and

$$\|f\|_{L_p(\mathcal{D})} = \left(\int_{\mathcal{D}} |f(x)|^p dx\right)^{1/p}.$$

We do not assume the intensity ϑ to be known: it is a nuisance parameter. We work under the following assumption.

Assumption 1.1. The jumps (ξ_i) have density f which is absolutely continuous with respect to the Lebesgue measure.

Encountered issues. By Assumption 1.1, in the presence of the event $\{X_{i\Delta} - X_{(i-1)\Delta} = 0\}$ no jump occurred between $(i-1)\Delta$ and $i\Delta$, thus the increment $X_{i\Delta} - X_{(i-1)\Delta}$ provides no information on f. When $\Delta \to 0$, many increments are zero, therefore to estimate f we focus on the nonzero ones and denote N_T their number over $[0, T]$. In this statistical context different difficulties arise. First, the sample size N_T is random. Second, in the presence of the event $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$ the increment $X_{i\Delta} - X_{(i-1)\Delta}$ is not necessarily a realization of the density f. Indeed, even if Δ is small, the probability that more than one jump occurred between $(i-1)\Delta$ and $i\Delta$ is positive. Conditional on $\{X_{i\Delta} - X_{(i-1)\Delta} \neq 0\}$, the law of $X_{i\Delta} - X_{(i-1)\Delta}$ has density given by (see Proposition 2.1)

$$\mathbf{P}_\Delta[f](x) = \sum_{m=1}^{\infty} \mathbb{P}\left(R_\Delta = m | R_\Delta \neq 0\right) f^{\star m}(x), \quad \text{for } x \in \mathbb{R}, \tag{3}$$

where $\star$ is the convolution product and $f^{\star m} = f \star \cdots \star f$, m times.

Estimation procedure. We build an estimator of f using (3). We proceed in two steps. First, we compute the inverse of the operator $f \to \mathbf{P}_\Delta[f]$. The inverse takes the form

$$\mathbf{P}_\Delta^{-1}[\nu] = \sum_{m \geq 1}^{\infty} a_m(\vartheta, \Delta_T)\nu^{\star m}, \quad \nu \in \mathcal{F}(\mathbb{R})$$

where $\mathcal{F}(\mathbb{R})$ denotes the space of densities with respect to the Lebesgue measure supported by $\mathbb{R}$ and the coefficients $\left(a_m(\vartheta, \Delta_T)\right)$ are explicit (see Proposition 2.1). They depend on the unknown intensity ϑ but can be estimated. This is referred to as decompounding in Buchmann and Grübel [6]. Then, we take advantage of

$$f \approx \mathbf{L}_{\Delta,K}\left[\mathbf{P}_\Delta[f]\right], \tag{4}$$

where $\mathbf{L}_{\Delta,K}$ is the Taylor expansion of order K in Δ of $\mathbf{P}_\Delta^{-1}$, which depends on $\big(\mathbf{P}_\Delta[f]^{\star m}, m = 1, \ldots, K+1\big)$.

Second, we estimate the densities $\mathbf{P}_\Delta[f]^{\star m}$, for $m = 1, \ldots, K+1$, from the nonzero increments which are independent with density $\mathbf{P}_\Delta[f]$. The main difficulty is that N_T is random. In Theorem 2.1 we show that conditional on N_T, wavelet threshold estimators of $\mathbf{P}_\Delta[f]^{\star m}$ attain a rate of convergence for the L_p loss (up to logarithmic factors which appear in Theorem 2.1) in $N_T^{-\alpha(s,\pi,p)}$, where $\alpha(s,p,\pi) \le 1/2$ depends on the regularity s, measured with the L_π norm, $\pi > 0$, of f (see *e.g.* Donoho et al. [14] and (17) hereafter). For T large enough we prove (see Proposition 5.1) that N_T concentrates around a deterministic value of the order of T, giving an unconditional rate of convergence in $T^{-\alpha(s,\pi,p)}$. Injecting these estimators into $\mathbf{L}_{\Delta,K}$, defined in (4), we obtain an estimator of f we call *estimator corrected at order* K.

Achievable rates.

Definition 1.1. An estimator $\widehat{f}$ of f built from the observations (1) is minimax (or optimal) on the class V for the L_p loss, $p \ge 1$ if

$$\sup_{f \in V} \mathbb{E}\big[\|\widehat{f} - f\|^p_{L_p(\mathcal{D})}\big] \simeq \inf_{\widetilde{f}} \sup_{f \in V} \mathbb{E}\big[\|\widetilde{f} - f\|^p_{L_p(\mathcal{D})}\big],$$

where the infimum is taken over all estimators.

To derive an upper bound for the L_p loss of the estimator corrected at order K we control two distinct error terms: a deterministic error in Δ_T^{K+1}, due to the approximation of f by $\mathbf{L}_{\Delta,K}\big[\mathbf{P}_\Delta[f]\big]$ in (4), and a statistical error in $T^{-\alpha(s,\pi,p)}$, due to the replacement of the $\mathbf{P}_\Delta[f]^{\star m}$ by estimators in the second step. In Theorem 2.1 we give an upper bound in (up to constants and inessential logarithmic factors which appear explicitly in Theorem 2.1)

$$\max\{T^{-\alpha(s,\pi,p)}, \Delta_T^{K+1}\}.$$

A lower bound in $T^{-\alpha(s,p,\pi)}$ is provided in Theorem 2.2. The upper bound decreases with K. Since $\alpha(s,\pi,p) \le 1/2$, if there exists K_0 such that

$$T\Delta_T^{2K_0+2} \le 1, \tag{5}$$

the estimator corrected at order K_0 attains minimax rates of convergence, in the sense of Definition 1.1, up to a logarithmic factor. For every Δ_T polynomially decreasing with T, it is possible to exhibit K_0 satisfying (5).

Remark 1.1. Actually, we provide an almost minimax estimator: the upper bound and the lower bound differ by an inessential logarithmic factor. At the expense of additional technicalities and in the particular case of L_2 loss functions, it is possible to remove the logarithmic term in the upper bound (see Cai [7] and Remark 2.3).

Remark 1.2. Adding correction terms when the sampling rate does not go fast enough to 0 appears in Kessler [26] to estimate in a parametric setting the drift and the diffusion coefficients of a diffusion from discrete observations. Kessler [26] exhibits an asymptotically efficient estimator provided the sampling rate satisfies (with our notation) $T\Delta^l \to 0$ as $T \to \infty$, for an arbitrary integer l. That condition improves on a former condition that imposed $T\Delta^2 \to 0$ (see *e.g.* Yoshida [38]).

Structure of the paper. The main results are given in Section 2; we define wavelet functions and Besov spaces, we construct our estimator corrected at order K and give the main Theorems 2.1 and 2.2. A numerical example illustrates its behavior in Section 3; we also compare numerically our estimator with an estimator defined in Comte and Genon-Catalot [10]. Finally, a discussion is provided in Section 4 and Section 5 is dedicated to the proofs.

2. Main results

2.1. Besov spaces and wavelet thresholding

We estimate the densities $\big(\mathbf{P}_{\Delta}[f]^{\star m}, m = 1, \ldots, K+1\big)$ using wavelet threshold density estimators and we study their performance uniformly over Besov balls. In this paragraph we reproduce some classical results on Besov spaces, wavelet bases and wavelet threshold estimators (see Cohen [9], Donoho et al. [14] or Kerkyacharian and Picard [25]) that are used in the next sections.

Wavelets and Besov spaces

Let (ϕ, ψ) be a pair of scaling function and mother wavelet which generate a regular wavelet basis adapted to the domain $\mathcal{D}$: for f in $L_p(\mathcal{D})$ we have

$$f = \sum_{k \in \Lambda_0} \alpha_{0k}\phi_{0k} + \sum_{j \geq 1}\sum_{k \in \Lambda_j} \beta_{jk}\psi_{jk}, \tag{6}$$

where $\phi_{0k}(\bullet) = \phi(\bullet - k)$, $\psi_{jk}(\bullet) = 2^{j/2}\psi(2^j \bullet - k)$,

$$\alpha_{0k} = \int \phi_{0k}(x) f(x)dx \quad \text{and} \quad \beta_{jk} = \int \psi_{jk}(x) f(x)dx.$$

For every $j \geq 0$, the set Λ_j has cardinality 2^j and incorporates boundary terms that we choose not to distinguish in notation for simplicity. We define Besov spaces in terms of wavelet coefficients: for $s > 0$ and $\pi \in (0, \infty]$ a function f belongs to the Besov space $\mathcal{B}^s_{\pi\infty}(\mathcal{D})$ if the norm

$$\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})} := \left(\sum_{k \in \Lambda_0} |\alpha_{0k}|^{\pi}\right)^{1/\pi} + \sup_{j \geq 0} 2^{j(s+1/2-1/\pi)}\left(\sum_{k \in \Lambda_j} |\beta_{jk}|^{\pi}\right)^{1/\pi} \tag{7}$$

is finite, with usual modifications if $\pi = \infty$. Additional properties on the wavelet basis generated by $\big(\phi, \psi\big)$ are needed; they are listed in the following assumption. To lighten notation in Assumption 2.1 (9) and (10), we introduce notation $(\varphi_\lambda, \lambda \in \Lambda)$ for the basis $(\phi_{0k}, \psi_{jk}, j \geq 1, k \in \Lambda_j)$ where φ stands for ϕ or ψ and λ concatenates indices j and k.

Assumption 2.1. For $p \geq 1$, we have the following.

- For some $\mathfrak{C} \geq 1$
$$\mathfrak{C}^{-1}2^{j(p/2-1)} \leq \|\psi_{jk}\|^p_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{j(p/2-1)}.$$
- For some $\mathfrak{C} > 0$, $\sigma > 0$ and for all $s \leq \sigma$, $J \geq 0$,

$$\left\| f - \sum_{k \in \Lambda_0} \alpha_{0k}\phi_{0k} - \sum_{j=1}^{J}\sum_{k \in \Lambda_j} \beta_{jk}\psi_{jk} \right\|_{L_p(\mathcal{D})} \leq \mathfrak{C}2^{-Js}\|f\|_{\mathcal{B}^s_{\pi\infty}(\mathcal{D})}. \tag{8}$$

- If $p \geq 1$, for some $\mathfrak{C} \geq 1$ and for any sequence of coefficients $(u_\lambda)_{\lambda\in\Lambda}$,

$$\mathfrak{C}^{-1}\Big\|\sum_{\lambda\in\Lambda} u_\lambda\varphi_\lambda\Big\|_{L_p(\mathcal{D})} \leq \Big\|\Big(\sum_{\lambda\in\Lambda}|u_\lambda\varphi_\lambda|^2\Big)^{1/2}\Big\|_{L_p(\mathcal{D})} \leq \mathfrak{C}\Big\|\sum_{\lambda\in\Lambda} u_\lambda\varphi_\lambda\Big\|_{L_p(\mathcal{D})}. \tag{9}$$

- For any subset $\Gamma \subset \Lambda$ and for some $\mathfrak{C} \geq 1$

$$\mathfrak{C}^{-1}\sum_{\lambda\in\Gamma}\|\varphi_\lambda\|^p_{L_p(\mathcal{D})} \leq \int_{\mathcal{D}}\Big(\sum_{\lambda\in\Gamma}|\varphi_\lambda(x)|^2\Big)^{p/2} \leq \mathfrak{C}\sum_{\lambda\in\Gamma}\|\varphi_\lambda\|^p_{L_p(\mathcal{D})}. \tag{10}$$

Property (8) ensures that definition (7) of Besov spaces matches the definition in terms of linear approximation. Property (9) ensures that $(\varphi_\lambda)_\lambda$ is an unconditional basis of L_p and (10) is a super-concentration inequality (see Kerkyacharian and Picard [25] p. 304 and p. 306).

Remark 2.1. Compactly supported wavelet bases satisfy Assumption 2.1 (see Kerkyacharian and Picard [25] 4.1.2 p. 304 and Theorem 4.2 p. 306).

Wavelet threshold estimator

Let (ϕ, ψ) be a pair of scaling function and mother wavelet that generates a basis satisfying Assumption 2.1 for some $\sigma > 0$. An estimator of a function f is obtained by replacing (α_{0k}) and (β_{jk}) in (6) by estimated values. We consider classical hard threshold estimators of the form

$$\widehat{f}(\bullet) = \sum_{k\in\Lambda_0}\widehat{\alpha_{0k}}\phi_{0k}(\bullet) + \sum_{j=1}^{J}\sum_{k\in\Lambda_j}\widehat{\beta_{jk}}\mathbb{1}_{\{|\widehat{\beta_{jk}}|\geq\eta\}}\psi_{jk}(\bullet),$$

where $\widehat{\alpha_{0k}}$ and $\widehat{\beta_{jk}}$ are estimators of α_{0k} and β_{jk}, and J and η are respectively the resolution level and the threshold, possibly depending on the data.

2.2. *Construction of the estimator*

Observations. Assume we observe X at times $i\Delta$ for some $\Delta > 0$

$$\boldsymbol{X} = \big(X_\Delta, \ldots, X_{\lfloor T\Delta^{-1}\rfloor\Delta}\big).$$

Introduce the increments $\mathbf{D}^\Delta X_i = X_{i\Delta} - X_{(i-1)\Delta}$, for $i = 1, \ldots, \lfloor T\Delta^{-1}\rfloor$, where $X_0 = 0$. They are independent and identically distributed since X is a compound Poisson process. Define

$$S_1 = \inf\{j, \mathbf{D}^\Delta X_j \neq 0\} \wedge \lfloor T\Delta^{-1}\rfloor$$
$$S_i = \inf\{j > S_{i-1}, \mathbf{D}^\Delta X_j \neq 0\} \wedge \lfloor T\Delta^{-1}\rfloor \quad \text{for } i \geq 1,$$

then S_i is the random index of the ith jump and

$$N_T = \sum_{i=1}^{\lfloor T\Delta^{-1}\rfloor}\mathbb{1}_{\{\mathbf{D}^\Delta X_i\neq 0\}}$$

is the number of observed nonzero increments over $[0, T]$. When $\Delta = \Delta_T \to 0$ as T goes to infinity, infinitely many increments are null and by Assumption 1.1 convey no information on f. Hence, we focus on the nonzero increments $\big(\mathbf{D}^\Delta X_{S_1}, \ldots, \mathbf{D}^\Delta X_{S_{N_T}}\big)$.

Proposition 2.1. *The distribution of the increment $\mathbf{D}^{\Delta} X_{S_1}$ has density with respect to the Lebesgue measure given by*

$$\mathbf{P}_{\Delta}[f] = \sum_{m=1}^{\infty} p_m(\Delta) f^{\star m},$$

where

$$p_m(\Delta) = \mathbb{P}\big(R_{\Delta} = m | R_{\Delta} \neq 0\big) = \frac{1}{e^{\vartheta \Delta} - 1} \frac{(\vartheta \Delta)^m}{m!}.$$

Let Δ_0 be such that

$$\sum_{m=2}^{\infty} \frac{(\vartheta \Delta_0)^{m-2}}{m!} \leq 1,$$

then for $\Delta \leq \Delta_0$, we have $1 - \vartheta \Delta \leq p_1(\Delta) \leq 1$.

Proof of Proposition 2.1 is given in the Appendix. It is straightforward that the nonlinear operator $\mathbf{P}_{\Delta}$ is a mapping from $\mathcal{F}(\mathbb{R})$ to itself. The observations $\big(\mathbf{D}^{\Delta} X_{S_i}\big)$ are realizations of the density $\mathbf{P}_{\Delta}[f]$ and by Proposition 2.1 we have $p_1(\Delta) \to 1$ as $\Delta = \Delta_T \to 0$. Then, for Δ_T small enough most of $\big(\mathbf{D}^{\Delta} X_{S_i}\big)$ have distribution f and an estimator of f may be any density estimator applied to $\big(\mathbf{D}^{\Delta} X_{S_i}\big)$. This is the approach of Shimizu [35] to estimate the Poissonian jump part of a Lévy process. That estimator requires a convergence condition on Δ_T to achieve minimax rate of convergence (see Shimizu [35], or Theorems 2.1 and 2.2).

Construction of the estimator corrected at order K. We adopt the strategy introduced in Section 1.2.

Lemma 2.1. *Let $\vartheta \Delta \leq \log(2)$, the inverse $\mathbf{P}_{\Delta}^{-1}$ of $\mathbf{P}_{\Delta}$, such that for all densities f in $\mathcal{F}(\mathbb{R})$ if $\mathbf{P}_{\Delta}[f] = \nu$ we have $\mathbf{P}_{\Delta}^{-1}[\nu] = f$, is given by*

$$\mathbf{P}_{\Delta}^{-1}[\nu] = \frac{1}{\vartheta \Delta} \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (e^{\vartheta \Delta} - 1)^m \nu^{\star m}.$$

Proof of Lemma 2.1 is given in the Appendix. Since $\mathbf{P}_{\Delta}^{-1}$ is a power series whose coefficients are equivalent to increasing powers of Δ, the Taylor expansion of order K in Δ of $\mathbf{P}_{\Delta}^{-1}$, denoted by $\mathbf{L}_{\Delta,K}$, is obtained by keeping the $K+1$ first terms of the inverse

$$\mathbf{L}_{\Delta,K}[\nu] = \frac{1}{\vartheta \Delta} \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} (e^{\vartheta \Delta} - 1)^m \nu^{\star m}, \quad \nu \in \mathcal{F}(\mathbb{R}). \tag{11}$$

Next, we build wavelet threshold density estimators of the $K+1$ first convolution powers of $\mathbf{P}_{\Delta}[f]$ that will be plugged in (11). Define for $m \geq 1$

$$\widehat{\alpha}_{0k}^{(m)} = \frac{1}{N_{T,m}} \sum_{i=1}^{N_{T,m}} \phi_{0k}\big(\mathbf{D}_m^{\Delta} X_{S_i}\big) \tag{12}$$

$$\widehat{\beta}_{0k}^{(m)} = \frac{1}{N_{T,m}} \sum_{i=1}^{N_{T,m}} \psi_{jk}\big(\mathbf{D}_m^{\Delta} X_{S_i}\big) \tag{13}$$

where $N_{T,m} = \lfloor N_T/m \rfloor \geq 1$ for large enough T (see Proposition 5.1) and for $i = 1, \ldots, N_{T,m}$

$$\mathbf{D}_m^{\Delta} X_{S_i} = \mathbf{D}^{\Delta} X_{S_{(i-1)m+1}} + \cdots + \mathbf{D}^{\Delta} X_{S_{im}}.$$

Remark 2.2. To get observations with density $\mathbf{P}_{\Delta}[f]^{\star m}$, we divide the dataset $\left(\mathbf{D}^{\Delta} X_{S_i}, i = 1, \ldots, N_T\right)$ in $N_{T,m}$ blocks of length m. By Lemma 5.2 hereafter and Proposition 2.1, those blocks are independent, and composed of independent variables distributed according to $\mathbf{P}_{\Delta}[f]$. Since we obtain the variables $(\mathbf{D}_m^{\Delta} X_{S_i}, i = 1, \ldots, N_{T,m})$ by adding all the variables in each block, those are independent and identically distributed with density $\mathbf{P}_{\Delta}[f]^{\star m}$.

Let $\eta > 0$ and $J \in \mathbb{N}$, define $\widehat{P_{\Delta,m}^{J,\eta}}$ the estimator of $\mathbf{P}_{\Delta}[f]^{\star m}$ over $\mathcal{D}$

$$\widehat{P_{\Delta,m}^{J,\eta}}(x) = \sum_k \widehat{\alpha}_{0k}^{(m)} \phi_{0k}(x) + \sum_{j=0}^{J} \sum_k \widehat{\beta}_{jk}^{(m)} \mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{(m)}| \geq \eta\right\}} \psi_{jk}(x), \quad x \in \mathcal{D}. \tag{14}$$

Definition 2.1. Let $\widetilde{f}_{T,\Delta}^{K}$ be the estimator corrected at order K, $K \geq 0$,

$$\widetilde{f}_{T,\Delta}^{K}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\left(e^{\widehat{\vartheta}_T \Delta} - 1\right)^m}{\widehat{\vartheta}_T \Delta} \widehat{P_{\Delta,m}^{J,\eta}}(x), \quad x \in \mathcal{D} \tag{15}$$

where

$$\widehat{\vartheta}_T = -\frac{1}{\Delta} \log(1 - \widehat{p}_T) \tag{16}$$

and $\widehat{p}_T = N_T / \lfloor T \Delta^{-1} \rfloor$ is the empirical estimator of $\mathbb{P}(R_{\Delta} \neq 0) = 1 - e^{-\vartheta \Delta}$.

Lemma 2.1 justifies the form of the estimator corrected at order K.

2.3. Convergence rates

We consider densities f satisfying a smoothness property in terms of Besov balls

$$\mathcal{F}(s, \pi, \mathfrak{M}) = \left\{ f \in \mathcal{F}(\mathbb{R}), \|f\|_{\mathcal{B}_{\pi\infty}^{s}(\mathcal{D})} \leq \mathfrak{M} \right\},$$

where $\mathfrak{M}$ is a positive constant. We estimate f on the compact interval $\mathcal{D}$, we only impose its restriction to $\mathcal{D}$ to belong to a Besov ball.

Theorem 2.1. *We work under Assumptions 1.1 and 2.1. Let $\sigma > s > 1/\pi$, $p \geq 1 \wedge \pi$ and $\widehat{P_{\Delta_T,m}^{J,\eta}}$ be the threshold wavelet estimator of $\mathbf{P}_{\Delta_T}[f]^{\star m}$ on $\mathcal{D}$ constructed from (ϕ, ψ) and defined in (14). Take J and η such that for some $\kappa > 0$*

$$2^J N_T^{-1} \log\left(N_T^{1/2}\right) \leq 1 \quad \text{and} \quad \eta = \kappa N_T^{-1/2} \sqrt{\log\left(N_T^{1/2}\right)}.$$

Let

$$\alpha(s, p, \pi) = \min\left\{ \frac{s}{2s+1}, \frac{s + 1/p - 1/\pi}{2\left(s + 1/2 - 1/\pi\right)} \right\}. \tag{17}$$

(1) *The estimator* $\widehat{P^{J,\eta}_{\Delta_T,m}}$ *satisfies for large enough* T *and sufficiently large* κ

$$\sup_{\mathbf{P}_{\Delta_T}[f]^{\star m}\in\mathcal{F}(s,\pi,\mathfrak{M})}\Big(\mathbb{E}\Big[\Big\|\widehat{P^{J,\eta}_{\Delta_T,m}}-\mathbf{P}_{\Delta_T}[f]^{\star m}\Big\|^p_{L_p(\mathcal{D})}|N_T\Big]\Big)^{1/p}$$
$$\leq \mathfrak{C}(\log(N_T))^{\mathfrak{c}}N_T^{-\alpha(s,p,\pi)}$$

where $\mathfrak{C}$ *depends on* $s,\pi,p,\mathfrak{M},\phi$ *and* ψ *and* $\mathfrak{c}$ *is defined as follows*

$$\mathfrak{c}=\begin{cases}\alpha(s,p,\pi), & \text{if } \pi\neq\frac{p}{(2s+1)}\\ \alpha(s,p,\pi)+1, & \text{otherwise.}\end{cases} \tag{18}$$

(2) *The estimator corrected at order* K *defined in* (15) *satisfies for* T *large enough and any positive constants* $\underline{\mathfrak{T}}$ *and* $\overline{\mathfrak{T}}$ $(\underline{\mathfrak{T}}<\overline{\mathfrak{T}})$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\|\widetilde{f}^K_{T,\Delta_T}-f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}\leq\mathfrak{C}\max\big\{(\log(T))^{\mathfrak{c}}T^{-\alpha(s,p,\pi)},\,\Delta_T^{K+1}\big\}$$

where $\mathfrak{C}$ *depends on* $s,\pi,p,\mathfrak{M},\phi,\psi,\underline{\mathfrak{T}},\overline{\mathfrak{T}}$ *and* K *and* $\mathfrak{c}$ *is defined in* (18).

Proof of Theorem 2.1 is postponed to Section 5. From a practical viewpoint, the sample size is N_T, which is why in Theorem 2.1 we give the resolution level J and the threshold η as functions of N_T instead of replacing N_T by its deterministic counterpart. An explicit bound for κ is given in Lemma 5.4.

Theorem 2.2. *Let* $\Delta_T\to 0$ *as* $T\to\infty$*, we have*

$$\underline{\lim}_{T\to\infty}\inf_{\widehat{f}}\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}T^{\alpha(s,p,\pi)}\big(\mathbb{E}\big[\|\widehat{f}-f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}>0,$$

where $\alpha(s,p,\pi)$ *is defined in* (17) *and the infimum is taken over all estimators built from the observations* (1).

Proof of Theorem 2.2 is postponed to Section 5. This lower bound is not surprising: without loss of generality assuming T is an integer, if we observe T independent realizations of f the minimax rate of convergence is in $T^{-\alpha(s,p,\pi)}$ (see for instance Donoho et al. [14] or Härdle et al. [21]). When X is continuously observed over $[0,T]$, there are exactly R_T independent realizations of f and for T large enough R_T and T are of the same order.

Theorem 2.1 ensures that the estimator corrected at order K attains the rate $T^{-\alpha(s,p,\pi)}$, which from Theorem 2.2 is minimax (in the sense of Definition 1.1, up to a logarithmic factor), for the smallest K such that

$$\Delta_T=O\big(T^{-\frac{\alpha(s,p,\pi)}{K+1}}\big).$$

Since $\alpha(s,p,\pi)\leq 1/2$ it is sufficient to choose K such that $T\Delta_T^{2K+2}=O(1)$. If Δ_T decays as a power of T *i.e.* if there exists $\delta>0$ such that for some $\mathfrak{C}>0$ we have $\Delta_T\leq\mathfrak{C}T^{-\delta}$, it is always possible to find a correction level K satisfying the previous constraint. The case $K=0$ corresponds to the approximation $f\approx\mathbf{P}_\Delta[f]$ (see Shimizu [35]), it has an upper bound in $\max\{T^{-\alpha(s,p,\pi)},\Delta_T\}$. It is minimax for instance if $T\Delta_T^2\leq 1$.

Remark 2.3. In the particular case of L_2 loss functions and for sampling rates satisfying $T\Delta^2\leq 1$, Figueroa-López [19], Shimizu [35], Comte and Genon-Catalot [10] and Bec and Lacour [4], provide upper bounds in $T^{-\alpha(s,p,\pi)}$, with no logarithmic term. In that case their procedures is better by a logarithmic term. But whenever Δ_T does not satisfy $T\Delta^2 \leq 1$ or for L_p loss functions, with $p \geq 1$ and $p \neq 2$, we give an almost minimax estimation procedure. In the particular case of L_2 loss functions it is possible to remove the logarithmic term in the upper bound (see Cai [7]).

3. Numerical implementation

A numerical example

We illustrate the behavior of the estimator corrected at order K when K increases. We compare its performance with an oracle: the wavelet estimator we would compute in the idealized framework where all the jumps are observed

$$\widehat{f}^{\mathrm{Oracle}}(x) = \sum_k \widehat{\alpha}_{0k}^{\mathrm{Oracle}} \phi_{0k}(x) + \sum_{j=0}^{J} \sum_k \widehat{\beta}_{jk}^{\mathrm{Oracle}} \mathbb{1}_{\left\{|\widehat{\beta}_{jk}^{\mathrm{Oracle}}| \geq \eta\right\}} \psi_{jk}(x),$$

$$\text{where } \widehat{\alpha}_{0k}^{\mathrm{Oracle}} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i) \quad \text{and} \quad \widehat{\beta}_{jk}^{\mathrm{Oracle}} = \frac{1}{R_T} \sum_{i=1}^{R_T} \phi_{0k}(\xi_i),$$

R_T being the value of the Poisson process R at time T and (ξ_i) the actual jumps. That estimator is used as a benchmark. It is the optimal estimator of the super experiment consisting in the continuous observation of the process X. Our procedure cannot perform as well as this oracle, however it is expected to give comparable results.

We estimate a compound Poisson process of intensity $\vartheta = 1$ on $[0, T]$ and of compound law given by the following mixture

$$f(x) = (1-a) f_1(x) + a f_2(x)$$

where f_1 is the density of a Gaussian $\mathcal{N}(0, 1)$ and f_2 of a Laplace with location parameter 1 and scale parameter 0.1. We take $a = 0.05$. We estimate f on $\mathcal{D} = [-6, 6]$ by the estimator corrected at order K, for different values of K, and by the Oracle. We choose the same parameters J and η and wavelet bases (ϕ, ψ) for the different estimators. We use the wavelet toolbox of Matlab and consider Symlets 4 wavelet functions and a resolution level $J = 10$. Symlets satisfy Assumption 2.1 since they are compactly supported (see Remark 2.1 and Härdle et al. [21] p. 66). We transform the data in an equispaced signal on a grid of length 2^L with $L = 8$, it is the binning procedure (see Härdle et al. [21] Chapter 12). The threshold is chosen as in Theorem 2.1. The estimators take the form of vectors giving the estimated values of f on the uniform grid $[-6, 6]$ with mesh 0.01.

Fig. 1 represents the corrected estimator for $K = 0$ and $K = 1$, the oracle and the true density f. All the estimators are computed on the same trajectory. They all manage to reproduce the shape of the density f. As expected the estimator corrected at order 1 seems better than the uncorrected one ($K = 0$). Moreover it gives a result comparable with the oracle estimator.

We measure the accuracy of the different estimators performing their L_2 errors. We approximate the L_2 errors by Monte Carlo, we compute $M = 1000$ times each estimator (for $T = 10\,000$ and $\Delta = 0.1$) and use the approximation

$$\mathbb{E}(\|\widehat{f} - f\|_2^2) \approx \frac{1}{M} \sum_{i=1}^{M} \left(\sum_{p=0}^{1200} \left(\widehat{f}(-6 + 0.01p) - f(-6 + 0.01p) \right)^2 \times 0.01 \right),$$

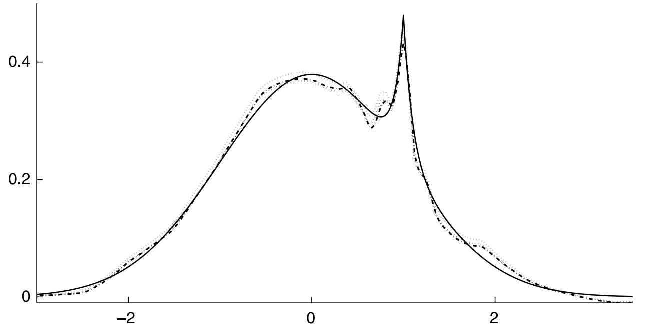

Fig. 1. Estimators of the density f for $T = 10\,000$ and $\Delta = 0.1$: true density (plain black), the uncorrected (dotted dark gray), the 1-corrected (dotted light gray) and the oracle (dashed dark).

where $\widehat{f}$ is one of the estimators and f the true density. For each Monte Carlo iteration the corrected estimators and the oracle are evaluated on the same trajectory. The results are reproduced in the following table.

Estimator	Oracle	$K = 0$	$K = 1$	$K = 2$	$K = 3$
L_2 error ($\times\, 10^{-2}$)	0.11	0.18	0.13	0.13	0.13
Standard deviation ($\times 10^{-3}$)	0.35	0.44	0.44	0.44	0.44

On this example there is an actual gain in considering the estimator corrected at order 1 instead of the uncorrected one. Indeed, here $p_1(\Delta) \approx 0.95$ which means that having no correction is equivalent to estimate f on a data set where 5% of the observations are realizations of a law which is not f. Considering more than 1 correction in this case is unnecessary, the L_2 losses stabilize afterwards. The L_2 loss of the oracle is strictly lower than the loss of the estimator corrected at order K, even for large K. That difference is explained by the fact that to estimate the mth convolution power we do not use N_T data points but $N_{T,m} = \lfloor N_T/m \rfloor$. Therefore we do not loose in terms of rate of convergence, but we surely deteriorate the constants in comparison with the oracle. These numerical results are consistent with the theoretical results of Theorem 2.1: in this example we took $T = 10\,000$ and $\Delta = 0.1$ thus $T\Delta^4 = 1$ which explains why here we do not observe improvements when correcting with K greater than 2.

Comparison with another estimator

The former example shows how the corrections manage to improve the estimation of the jump density. Here we compare, for different values of Δ, our estimator with an estimator of the Lévy density introduced in Comte and Genon-Catalot [10]. Their estimator, based on the Lévy–Khintchine formula, estimates the following function (with our notation and in the compound Poisson case)

$$g(x) = x\vartheta f(x),$$

where ϑ is the intensity and f the compound density. The estimator introduced in Comte and Genon-Catalot [10] denoted $\widehat{g}_m$ depends on a parameter m that can be chosen optimally. We compare the estimator $\widehat{g}_m$ of g with an estimator based on the estimator corrected at order K namely

$$\widetilde{g}_K = x\widehat{\vartheta}_T \widetilde{f}^{K}_{T,\Delta}(x).$$

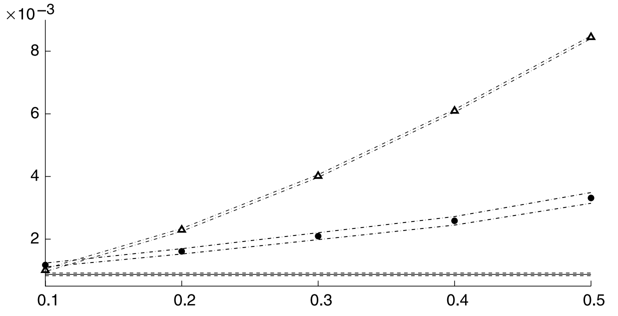

Fig. 2. The L_2 losses ($T = 10\,000$ and $M = 2000$) of the oracle $\widehat{g}^{\text{Oracle}}$ (plain line), $\widehat{g}_m$ (triangles) and $\widetilde{g}_K$ (points). Confidence intervals are given by dashed lines.

We investigate an example given in Comte and Genon-Catalot [10]: f is a centered Gaussian variable with variance 1 and $\vartheta = 1$. For this example the optimal choice of m is $\sqrt{\log(T)}/\pi$ (with our notation).

For $T = 10\,000$ and Δ in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$, we compute the estimator $\widehat{g}_m$ of [10] (using their parameters) and the estimator $\widetilde{g}_K$ where K is chosen such that $T\Delta^{2K+2} \simeq 1$: we have $K = 1$ for $\Delta = 0.1$, $K = 2$ for $\Delta = 0.2$, $K = 3$ for $\Delta = 0.3$, $K = 4$ for $\Delta = 0.4$ and $K = 6$ for $\Delta = 0.5$. Each estimator is computed on the uniform grid $[-4, 4]$ with mesh 0.01. We compare their L_2 errors for the different values of Δ. We also compute the L_2 error of the oracle $\widehat{g}^{\text{Oracle}} = x\vartheta\widehat{f}^{\text{Oracle}}(x)$, where $\widehat{f}^{\text{Oracle}}$ is defined above. To perform $\widetilde{g}_K$ and $\widehat{g}^{\text{Oracle}}$ we use Symlets 10 wavelet functions and a resolution level $J = 20$. We transform the data in an equispaced signal on a grid of length 2^L with $L = 8$. The threshold is chosen as in Theorem 2.1. The different L_2 losses are approximated by Monte Carlo using the former formula taking $M = 2000$. The results are reproduced in Fig. 2. For Δ sufficiently small ($\Delta = 0.1$) both estimators behave similarly. But when Δ increases our estimator $\widehat{g}_K$ appears to be better than the estimator $\widehat{g}_m$.

4. Discussion

4.1. Relation to other works

Compound Poisson process estimation and decompounding. The estimation problem considered here has been addressed by Buchmann and Grübel [6] and van Es et al. [37], for a fixed sampling rate $\Delta = 1$. In the discrete case, Buchmann and Grübel also invert the compounding operator to build their estimator. In the continuous case, they provide a formula giving the distribution function of the jumps F as a function of the distribution function of the increments G,

$$F(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1} e^{\lambda k}}{\lambda k} (G - G(0))^{\star k}(x).$$

They establish that replacing G by the empirical distribution function gives an asymptotically normal estimator for F. However, no bounds for the rate of convergence are given and from a practical viewpoint, no criterion indicates where the former sum should be truncated. In the continuous case, van Es et al. build their estimator using the Lévy–Khintchine formula. Assuming the intensity known, a consistent kernel density estimator is given, but no bounds ensuring minimaxity are derived.

Lévy processes estimation. On the larger class of Lévy processes, with the same sampling pattern and for L_2 loss functions, the nonparametric estimation of the Lévy density has been studied in great detail by Shimizu [35], Comte and Genon-Catalot [10,12], Figueroa-López [19] or Bec and Lacour [4]. Shimizu [35] estimates the Poisson jump part of a Lévy process; his estimator is minimax provided $T\Delta_T^2 \le 1$. Figueroa-López [19] builds a sieve estimator, which is minimax on Besov spaces for Δ's such that $T\Delta \le 1$. Comte and Genon-Catalot [10,12] construct a minimax adaptive nonparametric estimator of the Lévy density on Sobolev spaces and regimes satisfying $T\Delta_T \le 1$ ($T\Delta_T^2 \le 1$ under smoother assumptions). Bec and Lacour [4] obtain similar results when $T\Delta_T^2 \le 1$. The setting of Comte and Genon-Catalot [12] is more general; they estimate the Lévy triplet (drift, volatility and Lévy density) from a discretely observed Lévy process.

Our result is limited to the Poisson case contrary to the above references. However, we give a minimax procedure for L_p losses, $p \ge 1$, uniformly over Besov balls, for regimes where Δ_T is polynomially slow. When Δ_T decays even slower, for instance logarithmically in T, we still have a consistent estimator (see Theorem 2.1).

4.2. Extensions

Fixed sampling rates. If we now consider fixed sampling rates $\Delta > 0$, as studied by Buchmann Grübel [6] and van Es et al. [37] ($\Delta = 1$) or in the more general setting of Lévy processes as in Neumann and Reiß [32] or Comte and Genon-Catalot [11], it should be possible to generalize (to some extent) the procedure presented here. Let $\Delta_T \to \Delta_\infty < \log(2)/\vartheta$, the upper bound given in Theorem 2.1 should generalize in $\max\{T^{-\alpha(s,p,\pi)}, \Delta_\infty^{K+1}\}$. This time, the dependency in K of the constants needs to be controlled carefully since Δ no longer converges to 0.

Adding a continuous part. Generalizing our procedure to discretely observed Lévy processes with Poissonian jumps (*i.e.* finite Lévy measures) as in Shimizu [35], or more generally Comte and Genon-Catalot [12], seems delicate. A natural approach would be to threshold large observations to distinguish the jumps due to the Poisson process from the Brownian part (as in Shimizu [35]) and apply our procedure to the threshold observations. However, a threshold demands that Δ converges rapidly to 0 ($T\Delta^2 \le 1$ see *e.g.* Shimizu [35,36]), which is not the scope of this article. A methodology based on the Lévy–Khintchine formula (see Comte and Genon-Catalot [12]) imposes to work with L_2 losses and also demands that $T\Delta^2 \le 1$.

Shot noise processes. Another possible generalization would be to adapt our procedure to shot noise processes (see Mikosch [30])

$$Y_t = \sum_{i=1}^{R_t} \xi_i h(t - T_i), \quad t \ge 0$$

where R is a homogeneous Poisson process, (T_i) its jump times and h a deterministic function. These processes are used to model for instance electric currents (electrons arrive at times T_i with charge ξ_i and discharge according to h, see *e.g.* Mikosch [30]) or the claims of subscribers to insurance companies (h can be interpreted as a payoff, see *e.g.* Klüppelberg and Mikosch [27]). Our estimation procedure should adapt if applied to the threshold increments, to distinguish the jumps from the variations of h. However, depending on the jumps and/or the regularity of h, threshold conditions may not be consistent with the slow regimes of Δ considered here (see also the previous paragraph).

Other class of processes. We may consider more general jumps processes, like continuous time random walks, relaxing the memoryless property of compound Poisson processes. Such processes have many applications for instance in finance (see *e.g.* Cincotti et al. [8] or Meerschaert et al. [29]), insurance (see *e.g.* Mikosch [30]), biology (see *e.g.* Fedotov et al. [18] for tumor cells proliferation) or for earthquake modeling (see *e.g.* Helmstetter et al. [22]).

5. Proofs

In the sequel $\mathfrak{C}$ denotes a generic constant which may vary from line to line. Its dependencies may be indicated in the index.

5.1. Proof of part (1) of Theorem 2.1

To prove part (1) of Theorem 2.1 we apply the results of Kerkyacharian and Picard [25]. First, we establish some technical lemmas.

Preliminary lemmas

Lemma 5.1. *For any f in $\mathcal{F}(s,\pi,\mathfrak{M})$, $\mathbf{P}_\Delta[f]^{\star m}$ belongs to $\mathcal{F}(s,\pi,\mathfrak{M})$, $m\geq 1$.*

Proof of Lemma 5.1. It is straightforward to derive $\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{L_1(\mathbb{R})}=1$. The remainder of the proof is a consequence of the following statement. Let $f\in\mathcal{B}^s_{\pi\infty}$ and $g\in L_1$ we have

$$\|f\star g\|_{s\pi\infty}\leq\|f\|_{s\pi\infty}\|g\|_{L_1(\mathbb{R})}. \tag{19}$$

To prove (19), consider the following norm, equivalent to the Besov norm (see [21]),

$$\|\nu\|_{s\pi\infty}=\|\nu\|_{L_\pi(\mathbb{R})}+\|\nu^{(n)}\|_{L_\pi(\mathbb{R})}+\left\|\frac{w_\pi^2(\nu^{(n)},t)}{t^a}\right\|_\infty \tag{20}$$

where $s=n+a$, $n\in\mathbb{N}$, $a\in(0,1]$ and w_π is the modulus of continuity

$$w_\pi^2(\nu,t)=\sup_{|h|\leq t}\left\|\mathbf{D}^h\mathbf{D}^h[\nu]\right\|_{L_\pi(\mathbb{R})},$$

with $\mathbf{D}^h[\nu](x)=\nu(x-h)-\nu(x)$. Young's inequality gives

$$\|f_1\star f_2\|_{L_\pi(\mathbb{R})}\leq\|f_1\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}, \tag{21}$$

and the differentiation and translation invariance properties of the convolution product lead to, for $n\geq 1$

$$\left\|\frac{d^n}{dx^n}(f_1\star f_2)\right\|_{L_\pi(\mathbb{R})}=\left\|\left(\frac{d^n}{dx^n}f_1\right)\star f_2\right\|_{L_\pi(\mathbb{R})}\leq\left\|\frac{d^n}{dx^n}f_1\right\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})} \tag{22}$$

$$\text{and}\quad \begin{aligned}\left\|\mathbf{D}^h\mathbf{D}^h[(f_1\star f_2)^{(n)}]\right\|_{L_\pi(\mathbb{R})}&=\left\|(\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}])\star f_2\right\|_{L_\pi(\mathbb{R})}\\&\leq\left\|\mathbf{D}^h\mathbf{D}^h[f_1^{(n)}]\right\|_{L_\pi(\mathbb{R})}\|f_2\|_{L_1(\mathbb{R})}.\end{aligned} \tag{23}$$

Inequality (19) is then obtained by bounding (20) with (21), (22) and (23). Then, we apply $m-1$ times (19) which leads to

$$\forall m\in\mathbb{N}\setminus\{0\},\qquad\left\|\mathbf{P}_\Delta[f]^{\star m}\right\|_{s\pi\infty}\leq\left\|\mathbf{P}_\Delta[f]\right\|_{s\pi\infty}.$$

Finally, the triangle inequality gives $\|\mathbf{P}_\Delta[f]^{\star m}\|_{s\pi\infty}\leq\|f\|_{s\pi\infty}\leq\mathfrak{M}$ which completes the proof. $\square$

Lemma 5.2. *Let $n \geq 1$ and $\Delta > 0$. Then $(\mathbf{D}^{\Delta} X_{S_1}, \dots, \mathbf{D}^{\Delta} X_{S_n})$ are independent and identically distributed and independent of N_T.*

Proof of Lemma 5.2. The proof is the same as for the proof of the reject algorithm. The result is a consequence of the fact that

$$(S_i, i = 1, \dots, n) \text{ and } (\mathbf{D}^{\Delta} X_{S_i}, i = 1 \dots, n) \text{ are independent.} \tag{24}$$

If (24) holds, since X is a compound Poisson process $(\mathbf{D}^{\Delta} X_{S_1}, \dots, \mathbf{D}^{\Delta} X_{S_n})$ are independent and identically distributed. The independence between $(\mathbf{D}^{\Delta} X_{S_1}, \dots, \mathbf{D}^{\Delta} X_{S_n})$ and N_T is also derived from (24) as

$$N_T = \sum_{i=1}^{\infty} \mathbb{1}_{\{S_i < \lfloor T\Delta^{-1} \rfloor\}}.$$

The times (S_i) are stopping times for the filtration $\mathcal{F}_n = \sigma\big(\mathbf{D}^{\Delta} X_i, i = 1, \dots, n\big)$. Then the strong Markov property ensures the independence of the couples $(S_i, \mathbf{D}^{\Delta} X_{S_i})$. Moreover, since X is a compound Poisson process, for any measurable function h we get

$$\begin{aligned}
\mathbb{E}\big[h(\mathbf{D}^{\Delta} X_{S_i}) \mathbb{1}_{\{S_i = n\}}\big] &= \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_n) \mathbb{1}_{\{\mathbf{D}^{\Delta} X_{S_{i-1}+1} = 0, \dots, \mathbf{D}^{\Delta} X_{n-1} = 0, \mathbf{D}^{\Delta} X_n \neq 0\}}\big] \\
&= \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1) \mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big] \mathbb{P}(\mathbf{D}^{\Delta} X_1 = 0)^{n-1}
\end{aligned} \tag{25}$$

$$\begin{aligned}
\text{and} \quad \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_{S_i})\big] &= \sum_{k=1}^{\infty} \mathbb{P}(\mathbf{D}^{\Delta} X_1 = 0)^{k-1} \mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1) \mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big] \\
&= \frac{\mathbb{E}\big[h(\mathbf{D}^{\Delta} X_1) \mathbb{1}_{\{\mathbf{D}^{\Delta} X_1 \neq 0\}}\big]}{\mathbb{P}(\mathbf{D}^{\Delta} X_1 \neq 0)}.
\end{aligned} \tag{26}$$

Finally, as S_i is geometrically distributed with parameter $\mathbb{P}(\mathbf{D}^{\Delta} X_1 \neq 0)$, using (25) and (26) we derive (24). It concludes the proof. □

In the sequel we use (γ_{jk}) to design either (α_{0k}) or (β_{jk}) and (g_{jk}) for the wavelet functions (ϕ_{0k}) or (ψ_{jk}).

Lemma 5.3. *Let $2^j \leq N_T$. Then for all $m \in \mathbb{N} \setminus \{0\}$ and for $p \geq 1$ we have*

$$\mathbb{E}\big[\big|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\big|^p | N_T\big] \leq \mathfrak{C} N_T^{-p/2},$$

where $\mathfrak{C}$ depends on $p, m, \|g\|_{L_p(\mathbb{R})}, \mathfrak{M}$ and ϑ and $\widehat{\gamma}_{jk}^{(m)}$ is defined in (12) or (13) and

$$\gamma_{jk}^{(m)} = \int g_{jk}(y) \mathbf{P}_{\Delta}[f]^{\star m}(y) dy. \tag{27}$$

Proof of Lemma 5.3. The proof is obtained with Rosenthal's inequality: let $p \geq 1$ and let $(Y_1, \dots, Y_n)$ be independent random variables such that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\big[|Y_i|^p\big] < \infty$. Then there exists $\mathfrak{C}_p$ such that

$$\mathbb{E}\left[\left|\sum_{i=1}^{n} Y_i\right|^p\right] \leq \mathfrak{C}_p \left\{\sum_{i=1}^{n} \mathbb{E}\big[|Y_i|^p\big] + \left(\sum_{i=1}^{n} \mathbb{E}\big[|Y_i|^2\big]\right)^{p/2}\right\}. \tag{28}$$

The data $\left(\mathbf{D}_m^{\Delta_T} X_{S_i}\right)$ are independent and identically distributed with density $\mathbf{P}_{\Delta_T}[f]^{\star m}$ and $\mathbb{E}[\widehat{\gamma}_{jk}^{(m)}] = \gamma_{jk}^{(m)}$. Then $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is a sum of $N_{T,m} = \lfloor N_T/m \rfloor$ centered, independent and identically distributed random variables. Moreover

$$\begin{aligned}\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] &\le 2^p 2^{jp/2} \int |g(2^j y - k)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}(y) dy \\ &= 2^p 2^{j(p/2-1)} \int |g(z)|^p \mathbf{P}_{\Delta_T}[f]^{\star m}\Big(\frac{z+k}{2^j}\Big) dz \\ &\le 2^p 2^{j(p/2-1)} \|g\|_{L_p(\mathbb{R})}^p \big\|\mathbf{P}_{\Delta_T}[f]^{\star m}\big\|_\infty,\end{aligned}$$

where we made the substitution $z = 2^j x - k$. To control $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty$ we use the Sobolev embeddings (see [9,14,21])

$$\mathcal{B}_{\pi\infty}^s \hookrightarrow \mathcal{B}_{p\infty}^{s'} \quad \text{and} \quad \mathcal{B}_{\pi\infty}^{s'} \hookrightarrow \mathcal{B}_{\infty\infty}^s, \tag{29}$$

where $p > \pi, s\pi > 1$ and $s' = s - 1/\pi + 1/p$. It follows that $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \le \mathfrak{C}_{s,\pi} \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{\mathcal{B}_{\pi\infty(\mathcal{D})}^s}$, and Lemma 5.1 ensures $\|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_\infty \le \mathfrak{C}_{s,\pi}\mathfrak{M}$. We get for $p \ge 1$

$$\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^p\big] \le 2^p 2^{j(p/2-1)} \|g\|_{L_p(\mathbb{R})}^p \mathfrak{M}$$

and $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^2\big] \le \mathfrak{M}$ since $\|g\|_2^2 = 1$.

From Lemma 5.2, for all $n \ge 1$ the increments $\left(\mathbf{D}^{\Delta_T} X_{S_i}, i = 1, \ldots, n\right)$ are independent of N_T and thus of $N_{T,m}$. Therefore we apply Rosenthal's inequality conditional on N_T to $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ and derive for $p \ge 1$

$$\mathbb{E}\big[|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}|^p | N_T\big] \le \mathfrak{C}_p \left\{ 2^p \left(\frac{2^j}{N_{T,m}}\right)^{p/2-1} \|g\|_{L_p(\mathbb{R})}^p \mathfrak{M} + \mathfrak{M}^{p/2} \right\} N_{T,m}^{-p/2}.$$

The proof is complete. $\square$

Lemma 5.4. *Choose j and c such that*

$$2^j N_T^{-1} \log\big(N_T^{1/2}\big) \le 1 \quad \textit{and} \quad c^2 \ge \frac{16m}{3}\left(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\right).$$

For all $m \in \mathbb{N} \setminus \{0\}$ and $r \ge 1$, let $\kappa = \kappa_r = cr$. We have

$$\mathbb{P}\left(\left|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\right| \ge \frac{\kappa_r}{2} N_T^{-1/2} \sqrt{\log\big(N_T^{1/2}\big)} \middle| N_T\right) \le N_T^{-r/2},$$

where $\widehat{\gamma}_{jk}^{(m)}$ is defined in (12) or (13) and $\gamma_{jk}^{(m)}$ in (27).

Proof of Lemma 5.4. The proof is obtained with Bernstein's inequality. Let $Y_1, \ldots, Y_n$ be independent random variables such that $|Y_i| \le \mathfrak{A}, \mathbb{E}[Y_i] = 0$ and $b_n^2 = \sum_{i=1}^n \mathbb{E}[Y_i^2]$. Then for any $\lambda > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| > \lambda\right) \le 2\exp\left(-\frac{\lambda^2}{2\left(b_n^2 + \frac{\lambda\mathfrak{A}}{3}\right)}\right). \tag{30}$$

For all $m \geq 1$, $\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}$ is a sum of $N_{T,m}$ centered, independent and identically distributed random variables bounded by $2^{j/2}\|g\|_\infty$ and such that $\mathbb{E}\big[\big|g_{jk}(\mathbf{D}_m^{\Delta_T} X_{S_i})\big|^2\big] \leq \mathfrak{M}$. In view of Lemma 5.2 we apply Bernstein's inequality conditional on N_T

$$\mathbb{P}\left(\left|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\right| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\middle| N_T\right)$$
$$\leq 2\exp\left(-\frac{\kappa_r^2 N_T^{-1}\log\left(N_T^{1/2}\right)N_{T,m}^2}{8\left(N_{T,m}\mathfrak{M} + \frac{\kappa_r N_{T,m} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}2^{j/2}\|g\|_\infty}{6}\right)}\right)$$
$$= 2\exp\left(-\frac{c^2 r N_T^{-1} N_{T,m}}{8\left(\mathfrak{M} + \frac{\kappa_r N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}2^{j/2}\|g\|_\infty}{6}\right)} r\log\left(N_T^{1/2}\right)\right).$$

Using

$$m N_T^{-1} N_{T,m} = \frac{m}{N_T}\left\lfloor \frac{N_T}{m}\right\rfloor \geq \frac{3}{2},$$

for T large enough and $2^{j/2}N_T^{-1}\sqrt{\log\left(N_T^{1/2}\right)} \leq 1$ it follows that

$$\mathbb{P}\left(\left|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\right| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\middle| N_T\right)$$
$$\leq 2\exp\left(-\frac{3c^2 r}{16m\left(\mathfrak{M} + \frac{\kappa_r\|g\|_\infty}{6}\right)} r\log\left(N_T^{1/2}\right)\right).$$

With $c^2 \geq \frac{16m}{3}\left(\mathfrak{M} + \frac{c\|g\|_\infty}{6}\right)$ we get

$$\mathbb{P}\left(\left|\widehat{\gamma}_{jk}^{(m)} - \gamma_{jk}^{(m)}\right| \geq \frac{\kappa_r}{2} N_T^{-1/2}\sqrt{\log\left(N_T^{1/2}\right)}\middle| N_T\right) \leq N_T^{-r/2}.$$

The proof is complete. □

Completion of the proof of part (1) of Theorem 2.1

Part (1) of Theorem 2.1 is a consequence of Lemmas 5.1, 5.3, 5.4 and of the general theory of wavelet threshold estimators of [25]. It suffices to have conditions (5.1) and (5.2) of Theorem 5.1 of [25], which are satisfied – Lemmas 5.3 and 5.4 – with $c(T) = N_T^{-1/2}$ and $\Lambda_n = c(T)^{-1}$ (with notation of [25]). We apply Theorem 5.1, its Corollary 5.1 and Theorem 6.1 of [25] to obtain the result.

5.2. *Proof of part (2) of Theorem 2.1*

Part (1) of Theorem 2.1 is established conditional on N_T. To obtain part (2) of the theorem, we replace N_T by its deterministic counterpart using the following preliminary result.

Preliminary result

Proposition 5.1. *For all $r_1, r_2 \geq 0$, there exists $1 \leq \mathfrak{C}_\vartheta < \infty$, where $\vartheta \to \mathfrak{C}_\vartheta$ is continuous, such that*

$$1/\mathfrak{C}_\vartheta T^{-r_2} \leq \mathbb{E}\big[(\log(N_T))^{r_1} N_T^{-r_2}\big] \leq \mathfrak{C}_\vartheta (\log(T))^{r_1} T^{-r_2}.$$

Proof of Proposition 5.1. We have

$$N_T = \sum_{i=1}^{\lfloor T\Delta_T^{-1}\rfloor} \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}},$$

where we define

$$p(\Delta_T) = \mathbb{E}\big[\mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}}\big] = 1 - \exp(-\vartheta \Delta_T). \tag{31}$$

Introduce $(Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}} - p(\Delta_T))$, centered, independent and identically distributed random variables bounded by 2 and such that $\mathbb{E}[Y_i^2] \leq p(\Delta_T)$. The assumptions of Bernstein's inequality (30) are satisfied and we derive for $\lambda > 0$

$$\mathbb{P}\left(\left|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\right| > \lambda\right) \leq \exp\left(-\frac{\lfloor T\Delta_T^{-1}\rfloor \lambda^2}{2\left(p(\Delta_T) + \frac{2\lambda}{3}\right)}\right). \tag{32}$$

Setting $\lambda = p(\Delta_T)/2$, on the event $A_\lambda = \left\{\left|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\right| \leq \lambda\right\}$ we have

$$\lfloor T\Delta_T^{-1}\rfloor \frac{p(\Delta_T)}{2} \leq N_T \leq \lfloor T\Delta_T^{-1}\rfloor \frac{3p(\Delta_T)}{2}. \tag{33}$$

Moreover for Δ_T small enough we have from (31)

$$\frac{\vartheta \Delta_T}{2} \leq p(\Delta_T) \leq \vartheta \Delta_T. \tag{34}$$

The following decomposition holds

$$\mathbb{E}\big[(\log(N_T))^{r_1} N_T^{-r_2}\big] = \mathbb{E}\Big[(\log(N_T))^{r_1} N_T^{-r_2} \mathbb{1}_{\overline{\{A_\lambda\}}}\Big] + \mathbb{E}\Big[(\log(N_T))^{r_1} N_T^{-r_2} \mathbb{1}_{\{A_\lambda\}}\Big].$$

Since $r_1, r_2 \geq 0$ and $N_T \geq 1$, we obtain from inequalities (32)–(34) the bounds

$$\begin{aligned}
\mathbb{E}\big[(\log(N_T))^{r_1} N_T^{-r_2}\big] &\leq \left(\log\left(\frac{3\vartheta T}{2}\right)\right)^{r_1} \mathbb{P}\left(\left|\frac{N_T}{\lfloor T\Delta_T^{-1}\rfloor} - p(\Delta_T)\right| > \frac{p(\Delta_T)}{2}\right) \\
&\quad + \left(\log\left(\frac{3T\vartheta}{2}\right)\right)^{r_1} \left(\frac{T\vartheta}{4}\right)^{-r_2} \\
&\leq \left(\log\left(\frac{3\vartheta T}{2}\right)\right)^{r_1} \exp\left(-\frac{3\vartheta}{64} T\right) + \left(\log\left(\frac{3T\vartheta}{2}\right)\right)^{r_1} \left(\frac{T\vartheta}{4}\right)^{-r_2}
\end{aligned}$$

and

$$\mathbb{E}\big[(\log(N_T))^{r_1} N_T^{-r_2}\big] \geq \mathbb{E}\big[N_T^{-r_2}\big] \geq \left(\frac{3\lfloor T\Delta_T^{-1}\rfloor p(\Delta_T)}{2}\right)^{-r_2} \geq \left(\frac{3T\vartheta}{2}\right)^{-r_2}.$$

It follows that there exists $1 \le \mathfrak{C}_\vartheta < \infty$, where $\vartheta \to \mathfrak{C}_\vartheta$ is continuous, such that

$$1/\mathfrak{C}_\vartheta T^{-r_2} \le \mathbb{E}\big[(\log(N_T))^{r_1} N_T^{-r_2}\big] \le \mathfrak{C}_\vartheta \log(T)^{r_1} T^{-r_2}.$$

The proof is now complete. $\square$

Completion of the proof of part (2) of Theorem 2.1

Define for K in $\mathbb{N}$ and x in $\mathcal{D}$

$$\widehat{f}^K_{T,\Delta}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\left(e^{\vartheta\Delta} - 1\right)^m}{\vartheta\Delta} \widehat{P^{J,\eta}_{\Delta,m}}(x)$$

and decompose the L_p error as follows

$$\begin{aligned}\left(\mathbb{E}\big[\|\widetilde{f}^K_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\big]\right)^{1/p} &\le \left(\mathbb{E}\big[\|\widetilde{f}^K_{T,\Delta_T} - \widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]\right)^{1/p} \\ &\quad + \left(\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T} - f\|^p_{L_p(\mathcal{D})}\big]\right)^{1/p}.\end{aligned}$$

To complete the proof of Theorem 2.1 we bound each of these terms separately.

First, we get from the triangle inequality

$$\begin{aligned}&\left(\mathbb{E}\left[\left\|\sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{\vartheta\Delta_T} \widehat{P^{J,\eta}_{\Delta_T,m}} - \mathbf{P}^{-1}_{\Delta_T}\big[\mathbf{P}_{\Delta_T}[f]\big]\right\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \\ &\quad\le \sum_{m=1}^{K+1} \frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{m\vartheta\Delta_T} \left(\mathbb{E}\left[\left\|\widehat{P^{J,\eta}_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\right\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p} \qquad (35) \\ &\qquad + \sum_{m=K+2}^{\infty} \frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{m\vartheta\Delta_T} \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|_{L_p(\mathbb{R})}. \qquad (36)\end{aligned}$$

We bound (35) applying part (1) of Theorem 2.1, where the supremum is taken over the class $\{\mathbf{P}_{\Delta_T}[f]^{\star m} \in \mathcal{F}(s,\pi,\mathfrak{M})\}$. The inclusion

$$\{\mathbf{P}_{\Delta_T}[f]^{\star m}, f \in \mathcal{F}(s,\pi,\mathfrak{M})\} \subset \mathcal{F}(s,\pi,\mathfrak{M})$$

and Proposition 5.1 (applied with $r_1 = \alpha(s,p,\pi)$ and $r_2 = \alpha(s,p,\pi)$ if $\pi \ne p/(2s+1)$ or $r_2 = \alpha(s,p,\pi)+1$ otherwise) give for $m \ge 1$,

$$\mathbb{E}\left[\left\|\widehat{P^{J,\eta}_{\Delta_T,m}} - \mathbf{P}^{-1}_{\Delta_T}\big[\mathbf{P}_{\Delta_T}[f]\big]\right\|^p_{L_p(\mathcal{D})}\right] \le \mathfrak{C}(\log(T))^{\mathfrak{c}} T^{-\alpha(s,p,\pi)p}, \qquad (37)$$

where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, K, \vartheta$ and $\mathfrak{c}$ is defined in (18). We bound (36) using Young's inequality and $\left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_1(\mathbb{R})} = 1$ which give

$$\left\|\mathbf{P}_{\Delta_T}[f]^{\star m}\right\|_{L_p(\mathbb{R})} \le \left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_p(\mathbb{R})}, \quad \text{for } m \ge 1.$$

Then the triangle inequality leads to $\left\|\mathbf{P}_{\Delta_T}[f]\right\|_{L_p(\mathbb{R})} \le \|f\|_{L_p(\mathbb{R})}$ and Sobolev embeddings (29) to $\|f\|_{L_p(\mathbb{R})} \le \mathfrak{C}_{s,\pi,p}\mathfrak{M}$. Hence, we derive

$$\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{\vartheta\Delta_T}\left\|\mathbf{P}_{\Delta_T}[f]^{\star m}\right\|_{L_p(\mathbb{R})}\leq\|f\|_{L_p(\mathbb{R})}\sum_{m=K+2}^{\infty}\frac{1}{m}\frac{\left(e^{\vartheta\Delta_T}-1\right)^m}{\vartheta\Delta_T}$$
$$\leq\mathfrak{C}_{K,\vartheta,\mathfrak{M}}{\Delta_T}^{K+1}. \tag{38}$$

Finally, from (37) and (38) we obtain

$$\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\left(\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T}-f\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p}\leq\mathfrak{C}\max\left\{(\log(T))^{\mathfrak{c}}T^{-\alpha(s,p,\pi)},\,\Delta_T^{K+1}\right\},$$

where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi,K$ and ϑ. Since $\vartheta\to\mathfrak{C}$ is continuous we get

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\ \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\left(\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T}-f\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p}\leq\mathfrak{C}\max\left\{(\log(T))^{\mathfrak{c}}T^{-\alpha(s,p,\pi)},\,\Delta_T^{K+1}\right\}$$

where $\mathfrak{C}$ depends on $s,\pi,p,\mathfrak{M},\phi,\psi$ and K.

Second, we control $\mathbb{E}\left[\|\widetilde{f}^K_{T,\Delta_T}-\widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\right]$: from (16) we derive

$$\widetilde{f}^K_{T,\Delta_T}=\sum_{m=1}^{K+1}\frac{(-1)^m}{m}\frac{\left((1-\widehat{p}_T)^{-1}-1\right)^m}{\log(1-\widehat{p}_T)}\widehat{P^{J,\eta}_{\Delta_T,m}},$$

where $\widehat{P^{J,\eta}_{\Delta_T,m}}$ does not depend on ϑ (see (12) and (13)) and $\widehat{p}_T$ appears in Definition 2.1. Let

$$G_m(x)=\frac{((1-x)^{-1}-1)^m}{\log(1-x)}.$$

The triangle inequality leads to

$$\left(\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta})-\widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\right]\right)^{1/p}\leq\sum_{m=1}^{K+1}\Big(\mathbb{E}\Big[\|\big(G_m(\widehat{p}_T)$$
$$-G_m(p(\Delta_T))\big)\widehat{P^{J,\eta}_{\Delta_T,m}}\|^p_{L_p(\mathcal{D})}\Big]\Big)^{1/p},$$

where $p(\Delta_T)$ defined in (31) satisfies $p(\Delta_T)\leq\mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$ since $0<1-e^{-\underline{\mathfrak{T}}\Delta_T}\leq p(\Delta_T)\leq 1-e^{-\overline{\mathfrak{T}}\Delta_T}<1$. Moreover, we have for all $m\geq 1$

$$G'_m(x)=\frac{mx^{m-1}}{(1-x)^{m+1}\log(1-x)}+\frac{x^m}{(1-x)^{m+1}\left(\log(1-x)\right)^2}.$$

Then $xG'_m(x)$ is continuous over $(0,1/2]$ and goes to 0 as $x\to 0$. We deduce

$$\mathbb{E}\left[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta})-\widehat{f}^K_{T,\Delta_T}\|_{L_p(\mathcal{D})}\right]^{1/p}\leq\mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}},K}\Delta_T^{-1}\mathbb{E}\left[\left\|\left(\widehat{p}_T-p(\Delta_T)\right)\widehat{P^{J,\eta}_{\Delta_T,m}}\right\|^p_{L_p(\mathcal{D})}\right]^{1/p}.$$

The Cauchy–Schwarz inequality leads to

$$\mathbb{E}\left[\left\|\left(\widehat{p}_T-p(\Delta_T)\right)\widehat{P^{J,\eta}_{\Delta_T,m}}\right\|^p_{L_p(\mathcal{D})}\right]^2\leq\mathbb{E}\left[\left\|\left(\widehat{p}_T-p(\Delta_T)\right)\right\|^{2p}_{2p}\right]\mathbb{E}\left[\left\|\widehat{P^{J,\eta}_{\Delta_T,m}}\right\|^{2p}_{L_{2p}(\mathcal{D})}\right].$$

Then applying part (1) of Theorem 2.1 and as $N_T \geq 1$ we get

$$\mathbb{E}\Big[\|\widehat{P^{J,\eta}_{\Delta_T,m}}\|^{2p}_{L_{2p}(\mathcal{D})}\Big] \leq \mathbb{E}\Big[\|\widehat{P^{J,\eta}_{\Delta_T,m}} - \mathbf{P}_{\Delta_T}[f]^{\star m}\|^{2p}_{L_{2p}(\mathcal{D})}\Big] + \|\mathbf{P}_{\Delta_T}[f]^{\star m}\|^{2p}_{L_{2p}(\mathcal{D})}$$
$$\leq \mathfrak{C}\mathbb{E}[N_T^{-2\alpha(s,p,\pi)p}] + \mathfrak{M}^{2p} \leq \mathfrak{C} \tag{39}$$

where $\mathfrak{C}$ depends on $(s, \pi, p, \mathfrak{M}, \phi, \psi)$. We complete the proof with Rosenthal's inequality (28): $\widehat{p}_T - p(\Delta_T)$ is a sum of independent and identically distributed centered random variables

$$\big(Y_i = \mathbb{1}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}} - p(\Delta_T), i \in \{1, \ldots, \lfloor T\Delta_T^{-1}\rfloor\}\big)$$

where $\mathbb{E}[|Y_i|^{2p}] \leq 2^{2p}\mathbb{E}\big[\mathbb{1}^{2p}_{\{\mathbf{D}^{\Delta_T} X_i \neq 0\}}\big] \leq \mathfrak{C}_{p,\overline{\mathfrak{T}}}\Delta_T$ and $\mathbb{E}[|Y_i|^2] \leq \mathfrak{C}_{\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\Delta_T$. Rosenthal's inequality (28) leads to

$$\mathbb{E}\big[\|\widehat{p}_T - p(\Delta_T)\|^{2p}_{2p}\big] \leq \mathfrak{C}_{p,\underline{\mathfrak{T}},\overline{\mathfrak{T}}}\lfloor T\Delta_T^{-1}\rfloor^{-2p}\big(\lfloor T\Delta_T^{-1}\rfloor\Delta_T + (\lfloor T\Delta_T^{-1}\rfloor\Delta_T)^p\big). \tag{40}$$

Then from inequalities (39) and (40) we obtain

$$\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta}) - \widehat{f}^K_{T,\Delta_T}\|_{L_p(\mathcal{D})}\big]^{1/p} \leq \mathfrak{C}\Delta_T^{-1}\lfloor T\Delta_T^{-1}\rfloor^{-1}\big(T^{1/(2p)} + T^{1/2}\big),$$

where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{T}}, \overline{\mathfrak{T}}$ and K. We deduce for $p \geq 1$

$$\sup_{\vartheta\in[\underline{\mathfrak{T}},\overline{\mathfrak{T}}]}\ \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}\big[\|\widehat{f}^K_{T,\Delta_T}(\widehat{\vartheta}) - \widehat{f}^K_{T,\Delta_T}\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p} \leq \mathfrak{C}\big(T^{-(1-1/(2p))} + T^{-1/2}\big)$$

where $\mathfrak{C}$ depends on $s, \pi, p, \mathfrak{M}, \phi, \psi, \underline{\mathfrak{T}}, \overline{\mathfrak{T}}$ and K. This second term is negligible compared to $T^{-\alpha(s,p,\pi)}$ since $\alpha(s, p, \pi) \leq 1/2$. The proof of Theorem 2.1 is now complete.

5.3. *Proof of Theorem 2.2*

Step 1. *A super experiment*. If X is continuously observed over $[0, T]$, we observe $\widetilde{\boldsymbol{X}} = (X_s, s \in [0, T])$, and all R_T jumps (ξ_i) are recovered. We prove a lower bound in the super experiment, as taking the infimum over all estimators based on $\widetilde{\boldsymbol{X}}$ instead of $\boldsymbol{X}$ only reduces the lower bound. Denote $\mathbb{P}_{T,0}$ the law of $\widetilde{\boldsymbol{X}}$ and $\mathbb{P}_{T,\Delta}$ the law of $\boldsymbol{X}$. We have

$$\inf_{\widehat{f}}\ \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}_{\mathbb{P}_{T,\Delta}}\big[\|\widehat{f} - f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p}$$
$$\geq \inf_{\widetilde{f}}\ \sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\big(\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f} - f\|^p_{L_p(\mathcal{D})}\big]\big)^{1/p},$$

where the infimum on the left hand side is taken over all estimators built from $\boldsymbol{X}$ and on the right hand side over all estimators built from $\widetilde{\boldsymbol{X}}$.

Step 2. *A lower bound for the super experiment*. In the super experiment, there are R_T independent and identically distributed realizations of f, independent of R_T, since X is a compound Poisson process. Thus working conditional on R_T and applying classical lower bounds of the independent and identically distributed case (see for instance in Donoho et al. [14] or Härdle et al. [21]), we readily obtain for $k \in \mathbb{N}$

$$\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f} - f\|^p_{L_p(\mathcal{D})}|R_T = k\big]$$
$$\geq \int_{\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f} - f\|^p_{L_p(\mathcal{D})}|R_T = k\big]d\mu(f) \geq R_T^{-\alpha(s,p,\pi)},$$

where μ denotes a measure on the space $\mathcal{F}(s,\pi,\mathfrak{M})$. Taking expectation on R_T, whose distribution is denoted ν, we have with Fubini's theorem

$$\sum_{k=0}^{\infty}\nu(k)\int_{\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f}-f\|_{L_p(\mathcal{D})}^p|R_T=k\big]d\mu(f)$$
$$=\int_{\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f}-f\|_{L_p(\mathcal{D})}^p\big]d\mu(f).$$

It follows that

$$\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f}-f\|_{L_p(\mathcal{D})}^p\big]\geq\mathbb{E}\big[R_T^{-\alpha(s,p,\pi)}\big]$$
$$\geq\big(\mathfrak{C}_\vartheta T\big)^{-\alpha(s,p,\pi)}\mathbb{P}\Big(\Big|\frac{R_T}{T}-\vartheta\Big|\leq\sqrt{2\vartheta}\Big),$$

where $\mathfrak{C}_\vartheta=\vartheta+\sqrt{2\vartheta}$. The right hand part of the previous inequality can be bounded from below using Markov's inequality

$$\mathbb{E}\Bigg[\sup_{f\in\mathcal{F}(s,\pi,\mathfrak{M})}\mathbb{E}_{\mathbb{P}_{T,0}}\big[\|\widetilde{f}-f\|_{L_p(\mathcal{D})}^p|R_T\big]\Bigg]\geq\tfrac{1}{2}\big(\mathfrak{C}_\vartheta T\big)^{-\alpha(s,p,\pi)}.$$

Multiplying both sides by $T^{\alpha(s,p,\pi,)}$, taking the infimum over all estimators $\widetilde{f}$ and over the compact set Θ and taking the limit inferior in T complete the proof.

Acknowledgments

This work is a part of my Ph.D. Thesis under the supervision of Marc Hoffmann whom I would like to thank for his valuable remarks on this paper. My research is supported by a Ph.D. GIS Grant.

Appendix

A.1. Proof of Proposition 2.1

Let $x\in\mathbb{R}$, we have by stationarity of the increments of the process X

$$\mathbb{P}(\mathbf{D}^\Delta X_{S_1}\leq x)=\mathbb{P}(X_\Delta\leq x|X_\Delta\neq 0)=\sum_{m=1}^{\infty}p_m(\Delta)\mathbb{P}(X_\Delta\leq x|R_\Delta=m)$$

where for $m\geq 1$, $\mathbb{P}(X_\Delta\leq x|R_\Delta=m)=\int_{-\infty}^x f^{\star m}(y)dy$. Then, we readily obtain $\mathbb{P}(\mathbf{D}^\Delta X_{S_1}\leq x)=\int_{-\infty}^x\mathbf{P}_\Delta[f](y)dy$. Immediate computation gives $p_m(\Delta)$ for $m\geq 1$ and $p_1(\Delta)\leq 1$. Finally, since

$$\exp\big(\vartheta\Delta\big)-1=\vartheta\Delta\Bigg(1+\vartheta\Delta\sum_{m=2}^{\infty}\frac{\big(\vartheta\Delta\big)^{m-2}}{m!}\Bigg),$$

we derive that

$$g(\Delta):=\sum_{m=2}^{\infty}\frac{\big(\vartheta\Delta\big)^{m-2}}{m!}=\frac{1}{\big(\vartheta\Delta\big)^2}\big(\exp\big(\vartheta\Delta\big)-1-\vartheta\Delta\big)\longrightarrow\frac{1}{2}\quad\text{as }\Delta\to 0.$$

Since g is continuous, there exists $\Delta_0 > 0$ such that for all $\Delta \le \Delta_0$, $g(\Delta) \le 1$. Then $p_1(\Delta) \ge \frac{1}{1+\vartheta\Delta} \ge 1 - \vartheta\Delta$ follows.

A.2. *Proof of Lemma 2.1*

Let $\mathbf{F}[f]$ denote the Fourier transform of f and set $h = \mathbf{P}_\Delta[f]$. We use the one-to-one mapping between densities and their Fourier transforms to show Lemma 2.1. The linearity of the Fourier transform and relation $\mathbf{F}[f \star g] = \mathbf{F}[f]\mathbf{F}[g]$ give

$$\mathbf{F}[h] = \mathbf{F}\big[\mathbf{P}_\Delta[f]\big] = \frac{1}{e^{\vartheta\Delta} - 1} \sum_{m=1}^{\infty} \frac{\big(\vartheta\Delta\big)^m}{m!} \mathbf{F}[f]^m = \frac{\big(\exp(\vartheta\Delta\mathbf{F}[f]) - 1\big)}{e^{\vartheta\Delta} - 1},$$

from which we deduce

$$\mathbf{F}[f] = \frac{\log\big(1 + (e^{\vartheta\Delta} - 1)\mathbf{F}[h]\big)}{\vartheta\Delta} = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \frac{(e^{\vartheta\Delta} - 1)^m}{\vartheta\Delta} \mathbf{F}[h]^m$$

as $\big\|(e^{\vartheta\Delta} - 1)\mathbf{F}[h]\big\|_\infty < \big\|e^{\vartheta\Delta} - 1\big\|_\infty < 1$ holds for $\vartheta\Delta \le \log 2$. We obtain the result taking the inverse Fourier transform.

References

[1] H. Alexandersson, A simple stochastic mode of a precipitation process, Journal of Climate and Applied Meteorology 24 (1985) 1285–1295.

[2] M. Avellaneda, J. Reed, S. Stoikov, Forecasting prices in the presence of hidden liquidity, Preprint, 2010.

[3] M. Avellaneda, S. Stoikov, High frequency trading in a limit order book, Quantitative Finance 8 (2008) 271–283.

[4] M. Bec, C. Lacour, Adaptive kernel estimation of the Lévy density, Hal Preprint 00583221, 2012.

[5] R.J. Boys, D.J. Wilkinson, T.B.L. Kirkwood, Bayesian inference for a discretely observed stochastic kinetic model, Statistics and Computing 18 (2008) 125–135.

[6] B. Buchmann, R. Grübel, Decompounding: an estimation problem for Poisson random sums, The Annals of Statistics 31 (2003) 1054–1074.

[7] T. Cai, On block thresholding in wavelet regression: adaptivity, block size, and threshold level, Statistica Sinica 12 (2002) 1241–1273.

[8] S. Cincotti, L. Ponta, M. Raberto, E. Scalas, Poisson-process generalization for the trading waiting-time distribution in a double-auction mechanism, Working Paper, 2005.

[9] A. Cohen, Numerical Analysis of Wavelet Methods, Elsevier, 2003.

[10] F. Comte, V. Genon-Catalot, Nonparametric estimation for pure jump Lévy processes based on high frequency data, Stochastic Processes and their Applications 119 (2009) 4088–4123.

[11] F. Comte, V. Genon-Catalot, Nonparametric adaptive estimation for pure jump Lévy processes, Annales de l'Institut Henri Poincaré Probabilités et Statistiques 46 (2010) 595–617.

[12] F. Comte, V. Genon-Catalot, Estimation for Lévy processes from high frequency data within a long time interval, The Annals of Statistics 39 (2011) 803–837.

[13] R. Cont, A. de Larrard, Price dynamics in a Markovian limit order market, 2011, Preprint. arxiv:1104.4596v1.

[14] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, The Annals of Statistics 24 (1996) 508–539.

[15] C. Duval, M. Hoffmann, Statistical inference across time scales, Electronic Journal of Statistics 5 (2011) 2004–2030.

[16] P. Embrechts, C. Klüppelberg, M. Mikosch, Modelling Extremal Events, Springer, 1997.

[17] M.P. Etienne, E. Parent, H. Benoit, J. Bernier, Random effects compound Poisson model to represent data with extra zeros, 2009, Preprint. arxiv:0907.4903v1.

[18] S. Fedotov, A. Iomin, Probabilistic approach to a proliferation and migration dichotomy in the tumor cell invasion, 2008, Preprint. arxiv:0711.1304v2.

[19] J.E. Figueroa-López, Nonparametric estimation for Lévy models based on discrete sampling, in: Optimality: The Third Erich L. Lehmann Symposium, in: IMS Lecture Notes, Monogr. Ser., vol. 57, 2009, pp. 117–146.

[20] F. Guilbaud, H. Pham, Optimal high-frequency trading in a pro-rata microstructure with predictive information, 2012, Preprint. arxiv:1205.3051v1.
[21] W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, Approximation, and Statistical Applications, Springer-Verlag, New York, 1998.
[22] A. Helmstetter, D. Sornette, Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random walk models, Physical Review E 66 (2002) 061104.
[23] J.P. Huelsenbeck, B. Larget, D. Swofford, A compound Poisson process for relaxing the molecular clock, Genetics Society of America 154 (2000) 1879–1892.
[24] R.W. Katz, Stochastic modeling of Hurricane damage, Journal of Applied Meteorology 41 (2002) 754–762.
[25] G. Kerkyacharian, D. Picard, Thresholding algorithms, maxisets and well-concentrated bases, TEST 9 (2000) 283–344.
[26] M. Kessler, Estimation of an Ergodic diffusion from discrete observations, Scandinavian Journal of Statistics 24 (1997) 211–229.
[27] C. Klüppelberg, T. Mikosch, Explosive Poisson shot noise process with applications to risk reserves, Bernoulli 1 (1995) 125–147.
[28] J. Masoliver, M. Montero, J. Perelló, G.H. Weiss, Direct and inverse problems with some generalizations and extensions, 2008, Preprint. http://arxiv.org/abs/cond-mat/0308017.
[29] M.M. Meerschaert, E. Scalas, Coupled continuous time random walk in finance, Physica A 370 (2006) 114–118.
[30] T. Mikosch, Non-life Insurance Mathematics: An Introduction with the Poisson Process, Springer, 2009.
[31] P.S. Moharir, Estimation of the compounding distribution in the compound Poisson process model for earthquakes, Proceedings of the Indian Academy of Sciences 101 (1992) 347–359.
[32] M. Neumann, M. Reiß, Nonparametric estimation for Lévy processes from low-frequency observations, Bernoulli 15 (2009) 223–248.
[33] P. Repetowicz, M.M. Meerschaert, P. Richmond, Pricing of options on stocks that are driven by multi-dimensional coupled price-temporal infinitely divisible fluctuations, 2004, Preprint. http://arxiv.org/pdf/math-ph/0412071.pdf.
[34] E. Scalas, The application of continuous-time random walks in finance and economics, Physica A 362 (2006) 225–239.
[35] Y. Shimizu, Density estimation of Lévy measures for discretely observed diffusion processes with jumps, Journal of the Japan Statistical Society 36 (2006) 37–62.
[36] Y. Shimizu, Threshold selection in jump-discriminant filter for discretely observed jump processes, Statistical Methods and Applications 19 (2010) 355–378.
[37] B. van Es, S. Gugushvili, P. Spreij, A kernel type nonparametric density estimator for decompounding, Bernoulli 13 (2007) 672–694.
[38] N. Yoshida, Estimation for diffusion processes from discrete observations, Journal of Multivariate Analysis 41 (1992) 220–242.