Compound Poisson approximation to estimate the Lévy density

Céline Duval, Ester Mariucci

Compound Poisson approximation to estimate the Lévy density

Céline Duval [*] and Ester Mariucci [†]

Abstract

We construct an estimator of the Lévy density, with respect to the Lebesgue measure, of a pure jump Lévy process from high frequency observations: we observe one trajectory of the Lévy process over $[0, T]$ at the sampling rate Δ, where $\Delta \to 0$ as $T \to \infty$. The main novelty of our result is that we directly estimate the Lévy density in cases where the process may present infinite activity. Moreover, we study the risk of the estimator with respect to L_p loss functions, $1 \leq p < \infty$, whereas existing results only focus on $p \in \{2, \infty\}$. The main idea behind the estimation procedure that we propose is to use that "*every infinitely divisible distribution is the limit of a sequence of compound Poisson distributions*" (see e.g. Corollary 8.8 in Sato (1999)) and to take advantage of the fact that it is well known how to estimate the Lévy density of a compound Poisson process in the high frequency setting. We consider linear wavelet estimators and the performance of our procedure is studied in term of L_p loss functions, $p \geq 1$, over Besov balls. The results are illustrated on several examples.

Keywords. Density estimation, Infinite variation, Pure jump Lévy processes.
AMS Classification. 60E07, 60G51, 62G07, 62M99.

1 Introduction

Over the past decade, there has been a growing interest for Lévy processes. They are a fundamental building block in stochastic modeling of phenomena whose evolution in time exhibits sudden changes in value. Many of these models have been suggested and extensively studied in the area of *mathematical finance* (see e.g. [7] which explains the necessity of considering jumps when modeling asset returns). They play a central role in many other fields of science: in *physics*, for the study of turbulence, laser cooling and in quantum theory; in *engineering* for the study of networks, queues and dams; in *economics* for continuous time-series models, in *actuarial science* for the calculation of insurance and re-insurance risk (see e.g. [1, 4, 5, 32]).

[*]Sorbonne Paris Cité, Université Paris Descartes, MAP5, UMR CNRS 8145. E-mail: celine.duval@parisdescartes.fr

[†]Institut für Mathematik, Humboldt-Universität zu Berlin. E-mail: mariucce@math.hu-berlin.de. This research was supported by the Federal Ministry for Education and Research through the Sponsorship provided by the Alexander von Humboldt Foundation.

From a mathematical point of view the jump dynamics of a Lévy process X is dictated by its Lévy density. If it is continuous, its value at a point x_0 determines how frequent jumps of size close to x_0 are to occur per unit time. Thus, to understand the jump behavior of X, it is of crucial importance to estimate its Lévy density.

A problem that is now well understood is the estimation of the Lévy density of a compound Poisson process, that is, a pure jump Lévy process with a finite Lévy measure. There is a vast literature on the nonparametric estimation for compound Poisson processes both from high frequency and low frequency observations (see among others, [6, 8, 10, 14, 15], and [22] for the multidimensional setting).

Building an estimator of the Lévy density for a Lévy process X with infinite Lévy measure is a more demanding task; for instance, for any time interval $[0, t]$, the process X will almost certainly jump infinitely many times. In particular, its Lévy density, which we mean to estimate, is unbounded in any neighborhood of the origin. This implies that the techniques used for compound Poisson processes do not generalize immediately. The essential problem is that the knowledge that an increment $X_{t+\Delta} - X_t$ is larger than some $\varepsilon > 0$ does not give much insight on the size of the largest jump that has occurred between t and $t + \Delta$.

Many results are nevertheless already present in the literature concerning the estimation of the Lévy density from discrete data without the finiteness hypothesis, i.e., if we denote by f the Lévy density, when $\int_{\mathbb{R}} f(x)dx = \infty$. In that case, the main difficulty comes from the presence of small jumps and from the fact that the Lévy density blows up in a neighborhood of the origin. A number of different techniques has been employed to address this problem:

- To limit the estimation of f on a compact set away from 0;
- To study a functional of the Lévy density, such as $xf(x)$ or $x^2 f(x)$.

The analysis systematically relies on spectral approaches, based on the use of the Lévy-Khintchine formula (see (5) hereafter), that allows estimates for L_2 and L_∞ loss functions, but does not generalize easily to L_p for $p \notin \{2, \infty\}$. A non-exhaustive list of works related to this topic includes: [2, 9, 11, 12, 16, 18, 20, 21, 24, 25, 29, 35]; a review is also available in the textbook [3]. Projection estimators and their pointwise convergence has also been investigated in [17] and more recently in [27], where the maximal deviation of the estimator is examined. Two other works that, although not focused on constructing an estimator of f, are of interest for the study of Lévy processes with infinity activity in either low or high frequency are [30, 31]. Finally, from a theoretical point of view, one could use the asymptotic equivalence result in [28] to construct an estimator of the Lévy density f using an estimator of a functional of the drift in a Gaussian white noise model. However, any estimator resulting from this procedure would have the strong disadvantage of being randomized and, above all, would require the knowledge of the behavior of the Lévy density in a neighborhood of the origin.

The difference in purpose between the present work and the ones listed above is that we aim to build an estimator $\widehat{f}$ of f, without a smoothing treatment at the origin, and to study the following risk:

$$\mathbb{E}\bigg[\int_{A(\varepsilon)} |\widehat{f}(x) - f(x)|^p dx\bigg],$$

where, $\forall\varepsilon > 0$, $A(\varepsilon)$ is an interval away from 0: $A(\varepsilon)$ is included in $\mathbb{R}\setminus(-\varepsilon,\varepsilon)$ and such that $a(\varepsilon) := \min_{x\in A(\varepsilon)}|x| \geq \varepsilon$ where, possibly, $a(\varepsilon)\to 0$ as $\varepsilon\to 0$. To the knowledge of the authors the present work is the first attempt to build and study an estimator of the Lévy density on an interval that gets near the critical value 0 and whose risk is measured for L_p loss functions, $1 \leq p < \infty$.

More precisely, let X be a pure jump Lévy process with Lévy measure ν (see (3) for a precise definition) and suppose we observe

$$\big(X_{i\Delta} - X_{(i-1)\Delta},\ i = 1,\dots,n\big) \quad \text{with } \Delta\to 0 \text{ as } n\to\infty. \tag{1}$$

We want to estimate the density of the Lévy measure ν with respect to the Lebesgue measure, $f(x) := \frac{\nu(dx)}{dx}$, from the observations (1) on the set $A(\varepsilon)$ as $\varepsilon\to 0$. In this paper, the Lévy measure ν may have infinite variation, i.e.

$$\nu:\ \int_{\mathbb{R}}(x^2\wedge 1)\nu(dx) < \infty \ \text{ but possibly } \ \int_{|x|\leq 1}|x|\nu(dx) = \infty.$$

The starting point of this investigation is to look for a translation from a probabilistic to a statistical setting of Corollary 8.8 in [34]: "*Every infinitely divisible distribution is the limit of a sequence of compound Poisson distributions*". We are also motivated by the fact that a compound Poisson approximation has been successfully applied to approximate general pure jump Lévy processes, both theoretically and for applications. For example, using a sequence of compound Poisson processes is a standard way to simulate the trajectories of a pure jump Lévy process (see e.g. Chapter 6, Section 3 in [13]).

We briefly describe here the strategy of estimation for f on $A(\varepsilon)$. Given the observations (1), we choose $\varepsilon\in(0,1]$ (when $\nu(\mathbb{R}) < \infty$ the choice $\varepsilon = 0$ is also allowed) and we focus only on the observations such that $|X_{i\Delta} - X_{(i-1)\Delta}| > \varepsilon$. Let us denote by $\boldsymbol{n}(\varepsilon)$ the random number of observations satisfying this constraint. In the sequel, we informally mention the "small jumps" of the Lévy process X at time t when referring to one of the following objects. If ν is of finite variation, they are the sum of ΔX_s, the jumps of X at times $s\leq t$, that are smaller than ε in absolute value. If ν is of infinite variation, they correspond to the centered martingale at time t that is associated with the sum of the jumps the magnitude of which is less than ε in absolute value.

For Δ and ε small enough, the observations larger than ε in absolute value are in some sense close to the increments of a compound Poisson process $Z(\varepsilon)$ associated with the Lévy density $\mathbb{I}_{|x|>\varepsilon}\frac{\nu(dx)}{dx} =: \lambda_\varepsilon h_\varepsilon(x)$, where $\lambda_\varepsilon := \nu(\mathbb{R}\setminus(-\varepsilon,\varepsilon))$ and $h_\varepsilon(x) := \frac{1}{\lambda_\varepsilon}\frac{\nu(dx)}{dx}\mathbb{I}_{|x|>\varepsilon}$. It immediately follows that

$$f(x) = \lim_{\varepsilon\to 0}\lambda_\varepsilon h_\varepsilon(x)\mathbb{I}_{|x|>\varepsilon} \quad \forall x\neq 0.$$

Therefore, one can construct an estimator of f on $A(\varepsilon)$ by constructing estimators for λ_ε and h_ε separately. However, estimating h_ε and λ_ε from observed increments larger than ε is not straightforward due to the presence of the small jumps. In particular, if i_0 is such that $|X_{i_0\Delta} - X_{(i_0-1)\Delta}| > \varepsilon$, it is not automatically true that there exists $s\in((i_0-1)\Delta, i_0\Delta]$ such that $|\Delta X_s| > \varepsilon$ (or any other fixed positive number). Ignoring for a moment this difficulty and reasoning as if the increments of X larger than ε are increments of a compound Poisson process with intensity λ_ε and jumps density h_ε, the following estimators are constructed. First, a natural choice when estimating λ_ε is

$$\widehat{\lambda}_{n,\varepsilon} = \frac{\boldsymbol{n}(\varepsilon)}{n\Delta}. \tag{2}$$

In the special case where the Lévy measure ν is finite, we are allowed to take $\varepsilon = 0$ and the estimator (2), as $\Delta \to 0$, gets close to the maximum likelihood estimator. The study of the risk of this estimator with respect to an L_p norm is the subject of Theorems 1 and 2. Estimators of the cumulative distribution function of f have also been investigated in [30, 31], which is an estimation problem that is closely related to the estimation of λ_ε. However, as we detail in Section 3.2.2 below, the methodology used there does not apply in our setting: it cannot be adapted to L_p risks and the results of [30, 31] are established for non-vanishing ε whereas we are interested in having $\varepsilon \to 0$.

Second, for the estimation of the jump density h_ε, we fully exploit the fact that the observations are in high frequency. Under some additional constraints on the asymptotic of ε and Δ, we make the approximation $h_\varepsilon \approx \mathcal{L}(X_\Delta \big| |X_\Delta| > \varepsilon)$ and we apply a wavelet estimator to the increments larger than ε, in absolute value. The resulting estimator $\widehat{h}_{n,\varepsilon}$ is close to h_ε in L_p loss on $A(\varepsilon)$ (see Theorem 3 and Corollary 1). As mentioned above, the main difficulty in studying such an estimator is due to the presence of small jumps that are difficult to handle and limit the accuracy of the latter approximation. Also, we need to take into account that $\widehat{h}_{n,\varepsilon}$ is constructed from a random number of observations $\boldsymbol{n}(\varepsilon)$.

Finally, making use of the estimators $\widehat{\lambda}_{n,\varepsilon}$ and $\widehat{h}_{n,\varepsilon}$ we derive an estimator of f: $\widehat{f}_{n,\varepsilon} := \widehat{\lambda}_{n,\varepsilon}\widehat{h}_{n,\varepsilon}$ and study its risk in L_p norm. Our main result is then a consequence of Theorem 1 and Corollary 1 and is stated in Theorem 4.

It is easy to show that the upper bounds we provide tend to 0 (see Theorem 3, Corollary 1 and Theorem 4). It is also easy to check that in the particular cases where X is a compound Poisson process or a Gamma process, we recover usual results. Yet, it is tricky to compute the rate of convergence implied by these upper bounds in general. Indeed, the difficulty comes from the fact that for the estimation of both h_ε and λ_ε, quantities depending on the small jumps arise in the upper bounds. But even on simple examples, when the Lévy density is known, the distribution of the small jumps is unknown. We detail some examples where we can explicitly compute the rate of convergence of our estimation procedure. The obtained rates give evidence of the relevance of the approach presented here.

The paper is organized as follows. Preliminary Section 2 provides the statistical context as well as the necessary definitions and notations. An estimator of the intensity λ_ε is studied in Section 3.1 and a wavelet density estimator of the density h_ε in Section 3.3. Our main result is given in Section 4. Each of these results is illustrated on several examples. Finally Section 5 contains the proofs of the main theorems and an Appendix Section 6 collects the proofs of the auxiliary results.

2 Notations and preliminaries

We consider the class of pure jump Lévy processes with Lévy triplet $(\gamma_\nu, 0, \nu)$ where

$$\gamma_\nu := \begin{cases} \int_{|x|\leq 1} x\nu(dx) & \text{if} \quad \int_{|x|\leq 1} |x|\nu(dx) < \infty, \\ 0 & \text{otherwise.} \end{cases}$$

We recall that almost all paths of a pure jump Lévy process with Lévy measure ν have finite variation if and only if $\int_{|x|\leq 1} |x|\nu(dx) < \infty$. Thanks to the Lévy-Itô decomposition one can write a Lévy process X of Lévy triplet $(\gamma_\nu, 0, \nu)$ as the sum of two independent Lévy processes: for all $\varepsilon \in (0, 1]$

$$\begin{aligned} X_t &= tb_\nu(\varepsilon) + \lim_{\eta\to 0}\Big(\sum_{s\leq t} \Delta X_s \mathbb{1}_{(\eta,\varepsilon]}(|\Delta X_s|) - t\int_{\eta<|x|<\varepsilon} x\nu(dx)\Big) + \sum_{i=1}^{N_t(\varepsilon)} Y_i(\varepsilon) \\ &=: tb_\nu(\varepsilon) + M_t(\varepsilon) + Z_t(\varepsilon) \end{aligned} \tag{3}$$

where

- The drift $b_\nu(\varepsilon)$ is defined as

$$b_\nu(\varepsilon) := \begin{cases} \int_{|x|\leq \varepsilon} x\nu(dx) & \text{if} \quad \int_{|x|\leq 1} |x|\nu(dx) < \infty, \\ -\int_{\varepsilon\leq|x|<1} x\nu(dx) & \text{otherwise;} \end{cases} \tag{4}$$

- ΔX_r denotes the jump at time r of the càdlàg process X: $\Delta X_r = X_r - \lim_{s\uparrow r} X_s$;
- $M(\varepsilon) = (M_t(\varepsilon))_{t\geq 0}$ and $Z(\varepsilon) = (Z_t(\varepsilon))_{t\geq 0}$ are two independent Lévy processes of Lévy triplets $(0, 0, \mathbb{I}_{|x|\leq\varepsilon}\nu)$ and $(\int_{\varepsilon\leq|x|<1} x\nu(dx), 0, \mathbb{I}_{|x|>\varepsilon}\nu)$, respectively;
- $M(\varepsilon)$ is a centered martingale consisting of the sum of the jumps of magnitude smaller than ε in absolute value;
- $Z(\varepsilon)$ is a compound Poisson process defined as follows: $N(\varepsilon) = (N_t(\varepsilon))_{t\geq 0}$ is a Poisson process of intensity $\lambda_\varepsilon := \int_{|x|>\varepsilon} \nu(dx)$ and $(Y_i(\varepsilon))_{i\geq 1}$ are i.i.d. random variables independent of $N(\varepsilon)$ such that $\mathbb{P}(Y_1(\varepsilon) \in A) = \frac{\nu(A)}{\lambda_\varepsilon}$, for all $A \in \mathscr{B}(\mathbb{R} \setminus (-\varepsilon, \varepsilon))$.

The advantage of defining the drift $b_\nu(\varepsilon)$ as in (4) lies in the fact that this definition allows us to consider both the class of processes

$$X_t = \sum_{s\leq t} \Delta X_s,$$

when ν is of finite variation, and the class of Lévy processes with Lévy triplet $(0, 0, \nu)$, when ν is of infinite variation. Furthermore, when $\nu(\mathbb{R}) < \infty$, we can also take $\varepsilon = 0$ and then Equation (3) reduces to a compound Poisson process

$$X_t = \sum_{i=1}^{N_t} Y_i$$

where the intensity of the Poisson process is $\lambda = \lambda_0 = \nu(\mathbb{R} \setminus \{0\}) = \nu(\mathbb{R})$ and the density of the i.i.d. random variables $(Y_i)_{i\geq 0}$ is $f(x)/\lambda$.

We also recall that the characteristic function of any Lévy process X as in (3) can be expressed using the Lévy-Khintchine formula. For all u in $\mathbb{R}$, we have

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\bigg(itu\gamma_\nu + t\Big(\int_{\mathbb{R}} (e^{iuy} - 1 - iuy\mathbb{I}_{|y|\leq 1})\nu(dy)\Big)\bigg), \tag{5}$$

where ν is a measure on $\mathbb{R}$ satisfying

$$\nu(\{0\}) = 0 \quad \text{ and } \quad \int_{\mathbb{R}} (|y|^2 \wedge 1)\nu(dy) < \infty. \tag{6}$$

In the sequel we shall refer to $(\gamma^\nu, 0, \nu)$ as the *Lévy triplet* of the process X and to ν as the *Lévy measure*. This triplet characterizes the law of the process X uniquely.

Let us assume that the Lévy measure ν is absolutely continuous with respect to the Lebesgue measure and denote by f (resp. f_ε) the *Lévy density* of X (resp. $Z(\varepsilon)$), i.e. $f(x) = \frac{\nu(dx)}{dx}$ ($f_\varepsilon(x) = \frac{\mathbb{I}_{|x|>\varepsilon}\nu(dx)}{dx}$). Let h_ε be the density, with respect to the Lebesgue measure, of the random variables $(Y_i(\varepsilon))_{i\geq 0}$, i.e.

$$f_\varepsilon(x) = \lambda_\varepsilon h_\varepsilon(x)\mathbb{1}_{(\varepsilon,\infty)}(|x|).$$

We are interested in estimating f in any set of the form $A(\varepsilon) := (\overline{A}, -a(\varepsilon)] \cup [a(\varepsilon), \overline{A})$ where, for all $0 \leq \varepsilon \leq 1$, $a(\varepsilon)$ is a non-negative real number satisfying $a(\varepsilon) \geq \varepsilon$ and $\overline{A} \in [1, \infty]$ (the case $a(\varepsilon) = \varepsilon = 0$ is not excluded). The latter condition is technical, if X is a compound Poisson process we may choose $\overline{A} := +\infty$, otherwise we work under the simplifying assumption that $A(\varepsilon)$ is a bounded interval. Observe that, for all $A \subset A(\varepsilon)$ we have

$$f(x)\mathbb{1}_A(|x|) = \lambda_\varepsilon h_\varepsilon(x)\mathbb{1}_A(|x|). \tag{7}$$

In general, the Lévy density f goes to infinity as $x \downarrow 0$. It follows that if $a(\varepsilon) \downarrow 0$, for instance $a(\varepsilon) = \varepsilon$, we estimate a quantity that gets larger and larger. In the decomposition (7) of the Lévy density, the quantity that increases as ε goes to 0 is $\lambda_\varepsilon = \int_{|x|>\varepsilon} f(x)dx$, whereas the density $h_\varepsilon := \frac{f_\varepsilon}{\lambda_\varepsilon}$ may remain bounded in a neighborhood of the origin. The intensity λ_ε carries the information on the behavior of the Lévy measure f around 0.

Suppose we observe X on $[0, T]$ at the sampling rate $\Delta > 0$, without loss of generality, we set $T := n\Delta$ with $n \in \mathbb{N}$. Define

$$\boldsymbol{X}_{n,\Delta} := (X_\Delta, X_{2\Delta} - X_\Delta, \ldots, X_{n\Delta} - X_{(n-1)\Delta}). \tag{8}$$

We consider the high frequency setting where $\Delta \to 0$ and $T \to \infty$ as $n \to \infty$. The assumption $T \to \infty$ is necessary to construct a consistent estimator of f. To build an estimator of f on the interval $A(\varepsilon)$, we do not consider all the increments (8), but only those larger than ε in absolute value. Define the dataset $\boldsymbol{D}_{n,\varepsilon} := \{X_{i\Delta} - X_{(i-1)\Delta},\ i \in \mathscr{I}_\varepsilon\}$, where $\mathscr{I}_\varepsilon$ is the subset of indices such that

$$\mathscr{I}_\varepsilon := \{i = 1, \dots, n : |X_{(i-1)\Delta} - X_{i\Delta}| > \varepsilon\}.$$

Furthermore, denote by $\boldsymbol{n}(\varepsilon)$ the cardinality of $\mathscr{I}_\varepsilon$, i.e.:

$$\boldsymbol{n}(\varepsilon) := \sum_{i=1}^{n} \mathbb{1}_{\mathbb{R}\setminus[-\varepsilon,\varepsilon]}(|X_{i\Delta} - X_{(i-1)\Delta}|), \tag{9}$$

which is random.

We examine the properties of our estimation procedure in terms of L_p loss functions, restricted to the estimation interval $A(\varepsilon)$, for all $0 \leq \varepsilon \leq 1$. Let $p \geq 1$,

$$L_{p,\varepsilon} = \Big\{g : \|g\|_{L_{p,\varepsilon}} := \Big(\int_{A(\varepsilon)} |g(x)|^p dx\Big)^{\frac{1}{p}} < \infty\Big\}.$$

Define the loss function

$$\ell_{p,\varepsilon}(\widehat{f}, f) := \big(\mathbb{E}\big[\|\widehat{f} - f\|_{L_{p,\varepsilon}}^p\big]\big)^{1/p} = \Big(\int_{A(\varepsilon)} \mathbb{E}\big[|\widehat{f}(x) - f(x)|^p\big] dx\Big)^{1/p},$$

where $\widehat{f}$ is an estimator of f built from the observations $\boldsymbol{D}_{n,\varepsilon}$.

Finally, denote by P_Δ the distribution of the random variable X_Δ and by P_n the law of the random vector $\boldsymbol{X}_{n,\Delta}$ as defined in (8). Since X is a Lévy process, its increments are i.i.d., hence

$$P_n = \bigotimes_{i=1}^{n} P_{i,\Delta} = P_\Delta^{\otimes n}, \quad \text{where} \quad P_{i,\Delta} = \mathscr{L}(X_{i\Delta} - X_{(i-1)\Delta}).$$

We consider also the following family of product measures:

$$P_{n,\varepsilon} = \bigotimes_{i \in \mathscr{I}_\varepsilon} P_{i,\Delta}.$$

In the following, whenever confusion may arise, the reference probability in expectations is explicitly stated, for example, writing $\mathbb{E}_{P_n}$. The indicator function will be denoted equivalently as $\mathbb{1}_A(x)$ or $\mathbb{I}_{x\in A}$.

3 Main results

3.1 Estimation strategy

Contrary to existing results mentioned in Section 1, we adopt an estimation strategy that does not rely on the Lévy-Khintchine formula (5). A direct strategy based on projection estimators and their limiting distribution has been investigated in [17, 27]. Here, we consider a sequence of compound Poisson processes, indexed by $0 < \varepsilon \leq 1$, that gets close to X as $\varepsilon \downarrow 0$ and we approximate f using that

$$f(x) = \lim_{\varepsilon \searrow 0} \lambda_\varepsilon h_\varepsilon(x), \quad \forall x \in A(\varepsilon).$$

For each compound Poisson process $Z(\varepsilon)$, we build separately an estimator of its intensity, λ_ε, and a wavelet estimator for its jump density, h_ε. This leads to an estimator of $f_\varepsilon = \lambda_\varepsilon h_\varepsilon$. Therefore, we deal with two types of error; a deterministic approximation error arising when replacing f by f_ε and a stochastic error occurring when replacing f_ε by an estimator $\widehat{f}_{n,\varepsilon}$.

The main advantage of considering a sequence of compound Poisson processes is that, in the asymptotic $\Delta \to 0$, we can relate the density of the observed increments to the density of the jumps without going through the Lévy-Khintchine formula (see [14]). This approach enables the study of L_p loss functions, $1 \leq p < \infty$. Our estimation strategy is the following.

1. We build an estimator of λ_ε using the following result that is a minor modification of Lemma 6 in Rüschendorf and Woerner [33]. For sake of completeness we reproduce their argument in the Appendix.

   Lemma 1. *Let X be a Lévy process with Lévy measure ν. If g is a function such that $\int_{|x|\geq 1} g(x)\nu(dx) < \infty$, $\lim_{x\to 0} \frac{g(x)}{x^2} = 0$ and $\frac{g(x)}{(|x|^2 \wedge 1)}$ is bounded for all x in $\mathbb{R}$, then*

   $$\lim_{t\to 0} \frac{1}{t}\mathbb{E}[g(X_t)] = \int_{\mathbb{R}} g(x)\nu(dx).$$

   In particular, Lemma 1 applied to $g = \mathbb{1}_{A(\varepsilon)}$ implies that

   $$\lambda_\varepsilon = \lim_{\Delta\to 0} \frac{1}{\Delta}\mathbb{P}(|X_\Delta| > \varepsilon), \quad \forall\ 0 \leq \varepsilon \leq 1.$$

   Using this equation, we approximate λ_ε by $\frac{1}{\Delta}\mathbb{P}(|X_\Delta| > \varepsilon)$ and take the empirical counterpart of $\mathbb{P}(|X_\Delta| > \varepsilon)$.

2. From the observations $\boldsymbol{D}_{n,\varepsilon} = (X_{i\Delta} - X_{(i-1)\Delta})_{i\in\mathscr{I}_\varepsilon}$ we build a wavelet estimator $\widehat{h}_{n,\varepsilon}$ of h_ε relying on the approximation that for Δ small, the random variables $(X_{i\Delta} - X_{(i-1)\Delta})_{i\in\mathscr{I}_\varepsilon}$ are i.i.d. with a density close to h_ε (see Lemma 3).

3. Finally, we estimate f on $A(\varepsilon)$ following (7) by

   $$\widehat{f}_{n,\varepsilon}(x) := \widehat{\lambda}_{n,\varepsilon}\widehat{h}_{n,\varepsilon}(x)\mathbb{1}_{A(\varepsilon)}(|x|), \quad \forall x \in A(\varepsilon). \tag{10}$$

3.2 Statistical properties of $\widehat{\lambda}_{n,\varepsilon}$

3.2.1 Asymptotic and non-asymptotic results for $\widehat{\lambda}_{n,\varepsilon}$

First, we define the following estimator of the intensity of the Poisson process $Z(\varepsilon)$ in terms of $\boldsymbol{n}(\varepsilon)$, the number of jumps that exceed ε.

Definition 1. *Let* $\widehat{\lambda}_{n,\varepsilon}$ *be the estimator of* λ_ε *defined by*

$$\widehat{\lambda}_{n,\varepsilon} := \frac{\boldsymbol{n}(\varepsilon)}{n\Delta}, \tag{11}$$

where $\boldsymbol{n}(\varepsilon)$ *is defined as in* (9).

Controlling first the accuracy of the deterministic approximation of λ_ε by $\frac{1}{\Delta}\mathbb{P}(|X_\Delta| > \varepsilon)$ and second the statistical properties of the empirical estimator of $\mathbb{P}(|X_\Delta| > \varepsilon)$, we establish the following non-asymptotic bound for $\widehat{\lambda}_{n,\varepsilon}$.

Theorem 1. *For all* $n \geq 1$, $\Delta > 0$ *and* $\varepsilon \in [0,1]$, *let* $\widehat{\lambda}_{n,\varepsilon}$ *be the estimator of* λ_ε *introduced in Definition 1. Then, for all* $p \in [1,2)$, *we have*

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon} - \lambda_\varepsilon|^p\big] \leq C\Big\{\Big(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}} + \Big|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\Big|^p\Big\}$$

and, for all $p \in [2,\infty)$

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon} - \lambda_\varepsilon|^p\big] \leq C\Big\{\Big(\frac{n\mathbb{P}(|X_\Delta| > \varepsilon)}{(n\Delta)^p} \vee \Big(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}}\Big) + \Big|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\Big|^p\Big\},$$

where C *is a constant depending only on* p.

In the asymptotic setting the latter result simplifies as follows.

Theorem 2. *For all* $\Delta > 0$ *and* $\varepsilon \in [0,1]$, *let* $\widehat{\lambda}_{n,\varepsilon}$ *be the estimator of* λ_ε *introduced in Definition 1. Then, for all* $p \in [1,\infty)$, *we have*

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon} - \lambda_\varepsilon|^p\big] \leq \Big|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\Big|^p + O\Big(\Big(\frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}}\Big)$$

as $n \to \infty$, *provided that* $n\Delta$ *remains bounded away from 0 and that* $n\mathbb{P}(|X_\Delta| > \varepsilon) \to \infty$.

3.2.2 Some remarks on Theorems 1 and 2

On the convergence of $\widehat{\lambda}_{n,\varepsilon}$. Theorems 1 and 2 study how close is the estimator $\widehat{\lambda}_{n,\varepsilon}$ to the true value λ_ε, in L_p risk. The bound depends on the quantities $\mathbb{P}(|X_\Delta| > \varepsilon)$, which appears in the stochastic error, and $|\lambda_\varepsilon - \Delta^{-1}\mathbb{P}(|X_\Delta| > \varepsilon)|$, which represents the deterministic error of the estimator. Note that we may rewrite the stochastic error $|\widehat{\lambda}_{n,\varepsilon} - \frac{\mathbb{P}(|X_\Delta|>\varepsilon)}{\Delta}|$ as $\frac{1}{\Delta}|F_\Delta(\varepsilon) - \widehat{F}_{n,\Delta}(\varepsilon)|$, if we set $F_\Delta(\varepsilon) := \mathbb{P}(|X_\Delta| > \varepsilon)$ and $\widehat{F}_{n,\Delta}(\varepsilon)$ its empirical counterpart. Let us discuss what information we have on these terms.

If we decompose the stochastic error on the number $N_\Delta(\varepsilon)$ of jumps of the compound Poisson process $Z(\varepsilon)$ we get, for all $0 < \varepsilon \leq 1$,

$$\begin{aligned}\mathbb{P}(|X_\Delta| > \varepsilon) &\leq \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)| > \varepsilon)e^{-\lambda_\varepsilon\Delta} + u_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}\lambda_\varepsilon\Delta + \mathbb{P}(N_\Delta(\varepsilon) \geq 2)\\ &\leq v_\Delta(\varepsilon) + \lambda_\varepsilon\Delta + \frac{\lambda_\varepsilon^2\Delta^2}{2},\end{aligned}$$

where $v_\Delta(\varepsilon) := \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)| > \varepsilon)$ and $u_\Delta(\varepsilon) := \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon) + Y_1(\varepsilon)| > \varepsilon) \leq 1$. Note that, by Lemma 1, we have

$$\forall \varepsilon \in (0,1], \quad \lim_{\Delta \to 0} \frac{v_\Delta(\varepsilon)}{\Delta} = 0. \tag{12}$$

Indeed, the process $(M_t(\varepsilon) + tb_\nu(\varepsilon))_{t\geq 0}$ is a Lévy process with Lévy measure $\mathbb{1}_{[-\varepsilon,\varepsilon]}(x)\nu(dx)$. An application of Lemma 1 taking $g(x) = \mathbb{1}_{\mathbb{R}\setminus(-\varepsilon,\varepsilon)}$ gives us (12). This equation is important in the sequel: it gives an upper bound on the influence of the small jumps $M(\varepsilon)$.

For every fixed $\varepsilon \in (0,1]$, $F_\Delta(\varepsilon) = \mathbb{P}(|X_\Delta| > \varepsilon)$ is expected to converge to zero quickly enough as Δ goes to zero. Therefore, from the bound

$$\frac{1}{\Delta^p}\mathbb{E}_{P_n}\big[|F_\Delta(\varepsilon) - \widehat{F}_{n,\Delta}(\varepsilon)|^p\big] \leq C \begin{cases} \Big(\frac{v_\Delta(\varepsilon)}{n\Delta^2} + \frac{\lambda_\varepsilon^2}{n} + \frac{\lambda_\varepsilon}{n\Delta}\Big)^{\frac{p}{2}} & \text{if } p \in [1,2), \\ \frac{nF_\Delta(\varepsilon)}{(n\Delta)^p} \vee \Big(\frac{v_\Delta(\varepsilon)}{n\Delta^2} + \frac{\lambda_\varepsilon^2}{n} + \frac{\lambda_\varepsilon}{n\Delta}\Big)^{\frac{p}{2}} & \text{if } p \geq 2, \end{cases}$$

we deduce that $\lim_{n\to\infty} \frac{1}{\Delta^p}\mathbb{E}_{P_n}\big[|F_\Delta(\varepsilon) - \widehat{F}_{n,\Delta}(\varepsilon)|^p\big] = 0$ as long as we can choose ε such that both $\frac{\lambda_\varepsilon^2}{n}$ and $\frac{\lambda_\varepsilon}{n\Delta}$ vanish as n goes to infinity.

Let us now discuss the deterministic error term. We have

$$\begin{aligned} \Big|\lambda_\varepsilon - \frac{\mathbb{P}(|X_\Delta| > \varepsilon)}{\Delta}\Big| &= \Big|\lambda_\varepsilon - e^{-\lambda_\varepsilon \Delta}\Big(\tfrac{1}{\Delta}v_\Delta(\varepsilon) - u_\Delta(\varepsilon)\lambda_\varepsilon \\ &\qquad - \frac{1}{\Delta}\sum_{n=2}^{\infty} \mathbb{P}\Big(\Big|\sum_{i=1}^{n} Y_i + M_\Delta(\varepsilon) + b_\nu(\varepsilon)\Delta\Big| > \varepsilon\Big)\tfrac{(\lambda_\varepsilon\Delta)^n}{n!}\Big)\Big| \\ &\leq \frac{v_\Delta(\varepsilon)}{\Delta} + \lambda_\varepsilon(1 - u_\Delta(\varepsilon)) + \lambda_\varepsilon^2\Delta. \end{aligned}$$

For the term $\lambda_\varepsilon(1 - u_\Delta(\varepsilon))$, observe that, for all $\varepsilon, \varepsilon' \in [0,1]$

$$\begin{aligned} \lambda_\varepsilon(1 - u_\Delta(\varepsilon)) &\leq \lambda_\varepsilon\mathbb{P}(|Y_1| \leq \varepsilon + \varepsilon') + \lambda_\varepsilon\mathbb{P}(|Y_1| \geq \varepsilon + \varepsilon')v_\Delta(\varepsilon') \\ &\leq \nu\big([-\varepsilon - \varepsilon', -\varepsilon] \cup [\varepsilon, \varepsilon + \varepsilon']\big) + v_\Delta(\varepsilon')\lambda_\varepsilon. \end{aligned}$$

Therefore, if one chooses $\varepsilon' \in (0,1]$ such that

$$\begin{cases} v_\Delta(\varepsilon') \lesssim \frac{v_\Delta(\varepsilon)}{\Delta\lambda_\varepsilon}; \\ \nu\big([-\varepsilon - \varepsilon', -\varepsilon] \cup [\varepsilon, \varepsilon + \varepsilon']\big) \lesssim \frac{v_\Delta(\varepsilon)}{\Delta} \end{cases}$$

this term also goes to 0. Unfortunately, such a choice depends on the rate of converge in (12) which is very difficult to compute, even in examples. Therefore, it seems difficult to provide a general recipe for the choices of ε' and ε.

Relation to other works. In [30] and [31], the authors propose estimators of the cumulative distribution function of the Lévy measure, which is closely related to λ_ε. Indeed, following their notation, the authors estimate the quantity

$$\mathcal{N}(t) = \begin{cases} \int_{-\infty}^{t} \nu(dx), & \text{if } t < 0 \\ \int_{t}^{\infty} \nu(dx), & \text{if } t > 0. \end{cases}$$

Then, for all $\varepsilon \in (0,1]$, we have $\lambda_\varepsilon = \mathcal{N}(-\varepsilon) + \mathcal{N}(\varepsilon)$. The low frequency case is investigated in [30] ($\Delta > 0$) whereas [31] considers the high frequency setting ($\Delta \to 0$) and includes the possibility that the Brownian part is nonzero. In both cases, an estimator of $\mathcal{N}$ based on a spectral approach, relying on the Lévy-Khintchine formula (5), is studied. In [31] a direct approach equivalent to our estimator is also proposed and studied.

For each of these estimators the performances are investigated in L_∞ and functional central limit theorems are derived. However, the involved techniques use empirical processes and cannot be generalized for L_p losses, $p \geq 1$. Most importantly, those results hold for values of t that cannot get close to 0, whereas in our case we require an estimator at a time ε that is vanishing. Therefore, in this context, our Theorems 1 and 2 are new.

A corrected estimator. If we had a better understanding of the rate (12) we could improve the estimator $\widehat{\lambda}_{n,\varepsilon}$ in some cases. A trivial example is the case where X is a compound Poisson process. Then, one should set $\varepsilon = 0$, as we have exactly

$$\mathbb{P}(|X_\Delta| > 0) = \mathbb{P}(N_\Delta(0) \neq 0) = 1 - e^{-\lambda_0 \Delta}.$$

Replacing $\mathbb{P}(|X_\Delta| > 0)$ with its empirical counterpart $\widehat{F}_{n,\Delta}(0)$ and inverting the equation, one obtains an estimator of λ_0 converging at rate $\sqrt{n\Delta}$ (see e.g. [14]). A more interesting example is the case of subordinators, i.e. pure jump Lévy processes of finite variation and Lévy measure concentrated on $(0,\infty)$. If X is a subordinator of Lévy measure $\nu = \mathbb{1}_{(0,\infty)}\nu$, using the fact that $\mathbb{P}(Z_\Delta(\varepsilon) > \varepsilon | N_\Delta(\varepsilon) \neq 0) = 1$, we get

$$\mathbb{P}(X_\Delta > \varepsilon) = \mathbb{P}(M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon) > \varepsilon) e^{-\lambda_\varepsilon \Delta} + 1 - e^{-\lambda_\varepsilon \Delta}, \quad \varepsilon > 0. \tag{13}$$

Suppose we know additionally that

$$v_\Delta(\varepsilon) = \mathbb{P}(M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon) > \varepsilon) = o\big(F_\Delta(\varepsilon)^K\big) \tag{14}$$

for some integer K. Equation (12) as well as $F_\Delta(\varepsilon) = O(\lambda_\varepsilon \Delta)$ ensures that $K \geq 1$ (neglecting the influence of λ_ε with respect to Δ). Using the same notations as above, define the corrected estimator at order K

$$\widetilde{\lambda}_{n,\varepsilon}^{K} := \frac{1}{\Delta} \sum_{k=1}^{K} \frac{\big(\widehat{F}_{n,\Delta}(\varepsilon)\big)^k}{k}, \quad K \geq 1.$$

If $K=1$ we have $\widetilde{\lambda}^1_{n,\varepsilon}=\widehat{\lambda}_{n,\varepsilon}$. For $1\le p<\infty$, straightforward computations lead to

$$\mathbb{E}\big((\widetilde{\lambda}^K_{n,\varepsilon}-\lambda_\varepsilon)^p\big)\le C_p\Big\{\frac{1}{\Delta^p}\mathbb{E}\Big(\Big|\sum_{k=1}^K\frac{\big(\widehat{F}_{n,\Delta}(\varepsilon)\big)^k}{k}-\frac{\big(F_\Delta(\varepsilon)\big)^k}{k}\Big|^p\Big)+\Big|\frac{1}{\Delta}\sum_{k=1}^K\frac{\big(F_\Delta(\varepsilon)\big)^k}{k}-\lambda_\varepsilon\Big|^p\Big\}$$
$$\le C_p\Big\{C_{K,p}\frac{\mathbb{E}\big(|\widehat{F}_{n,\Delta}(\varepsilon)-F_\Delta(\varepsilon)|^p\big)}{\Delta^p}+\frac{1}{\Delta^p}\Big|\sum_{k=1}^K\frac{\big(F_\Delta(\varepsilon)\big)^k}{k}-\log\Big(\frac{1-v_\Delta(\varepsilon)}{1-F_\Delta(\varepsilon)}\Big)\Big|^p\Big\}$$

where we used (13). Finally, using the proof of Theorem 2, expansion at order K of $\log(1-x)$ in 0 and assumption (14) we easily derive

$$\mathbb{E}\big[(\widetilde{\lambda}^K_{n,\varepsilon}-\lambda_\varepsilon)^p\big]\le C\Big(\frac{F_\Delta(\varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}}\vee\Big(\frac{F_\Delta(\varepsilon)^{(K+1)p}}{\Delta^p}\Big).$$

However, even when the Lévy density is known, we do not know how to compute $\mathbb{P}(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon)>\varepsilon)$: assumption (14) is hardly tractable in practice. In the case of a subordinator, taking advantage of (13) and $\lambda_\varepsilon\Delta\to 0$ it is straightforward to have $v_\Delta(\varepsilon)=O\big(\Delta^{-1}F_\Delta(\varepsilon)-\lambda_\varepsilon\big)$, when $\lambda_\varepsilon\neq 0$. In many examples $\Delta^{-1}F_\Delta(\varepsilon)-\lambda_\varepsilon=O(\lambda_\varepsilon^2\Delta^2)$, therefore $v_\Delta(\varepsilon)=O(\lambda_\varepsilon^2\Delta^2)$. In these cases one should prefer the estimator $\widetilde{\lambda}^2_{n,\varepsilon}$.

3.2.3 Examples

Compound Poisson process. Let X be a compound Poisson process with Lévy measure ν and intensity λ (i.e. $0<\lambda=\nu(\mathbb{R})<\infty$). As ν is a finite Lévy measure, we take $\varepsilon=0$ in (11) that is,

$$\widehat{\lambda}_{n,0}=\frac{\sum_{i=1}^n\mathbb{1}_{\mathbb{R}\setminus\{0\}}(|X_{i\Delta}-X_{(i-1)\Delta}|)}{n\Delta}.$$

Applying Theorem 1 (and observing that $\lambda_0=\lambda$), we have the following result.

Proposition 1 (Compound Poisson Process)**.** *For all $n\ge 1$ and for all $\Delta>0$ such that $\lambda\Delta\le 1$, there exist constants C_1 and C_2, only depending on p, such that*

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,0}-\lambda|^p\big]\le C_1\Big\{\Big(\frac{\lambda}{n\Delta}\Big)^{\frac{p}{2}}+(\lambda^2\Delta)^p\Big\},\quad \text{if } p\in[1,2),$$
$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,0}-\lambda|^p\big]\le C_2\Big\{\frac{1}{(n\Delta)^{p-1}}\vee\Big(\frac{\lambda}{n\Delta}\Big)^{\frac{p}{2}}+(\lambda^2\Delta)^p\Big\},\quad \text{if } p\ge 2.$$

This rate depends on the rate at which Δ goes to 0, and the bound of Δ^p might, in some cases, be slower than the parametric rate in $(n\Delta)^{-p/2}=T^{-p/2}$. Indeed, the reason for this lies in the exact bound $|\lambda_0-\frac{\mathbb{P}(|X_\Delta|\neq 0)}{\Delta}|=|\lambda-(1-e^{-\lambda\Delta})|=O(\Delta)$. In the compound Poisson case another estimator of λ converging at parametric rate can be constructed using the Poisson structure of the problem (see e.g. [14], one may also use the corrected estimator discussed above).

Gamma process. Let X be a Gamma process of parameter $(1,1)$, that is a finite variation Lévy process with Lévy density $f(x)=\frac{e^{-x}}{x}\mathbb{1}_{(0,\infty)}(x)$, $\lambda_\varepsilon=\int_\varepsilon^\infty \frac{e^{-x}}{x}dx$ and

$$\mathbb{P}(|X_t|>\varepsilon)=\mathbb{P}(X_t>\varepsilon)=\int_\varepsilon^\infty \frac{x^{t-1}}{\Gamma(t)}e^{-x}dx,\quad \forall\varepsilon>0,$$

where $\Gamma(t)$ denotes the Γ function, i.e. $\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}dx$. By Theorem 1, an upper bound for $\mathbb{E}_{P_n}|\widehat{\lambda}_{n,\varepsilon}-\lambda_\varepsilon|^p$ can be expressed in terms of the quantities $\big|\lambda_\varepsilon-\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}\big|$ and $\mathbb{P}(X_\Delta>\varepsilon)$ that can be made explicit. Let us begin by computing the first term

$$\Big|\lambda_\varepsilon-\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}\Big|=\Big|\int_\varepsilon^\infty\frac{e^{-x}}{x}dx-\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}\Big|. \tag{15}$$

Define $\Gamma(\Delta,\varepsilon)=\int_\varepsilon^\infty x^{\Delta-1}e^{-x}dx$, such that $\Gamma(\Delta,0)=\Gamma(\Delta)$. Using that $\Gamma(\Delta,\varepsilon)$ is analytic we can write the right hand side of (15) as

$$\begin{aligned}\Big|\lambda_\varepsilon-\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}\Big|&=\frac{1}{\Delta\Gamma(\Delta)}\Big|\Delta\Gamma(\Delta,0)\Gamma(0,\varepsilon)-\sum_{k=0}^\infty\frac{\Delta^k}{k!}\Big\{\frac{\partial^k}{\partial\Delta^k}\Gamma(\Delta,\varepsilon)\Big|_{\Delta=0}\Big\}\Big|\\&\leq\Gamma(0,\varepsilon)\Big|\frac{1-\Delta\Gamma(\Delta,0)}{\Delta\Gamma(\Delta)}\Big|+\Big|\frac{1}{\Delta\Gamma(\Delta)}\sum_{k=1}^\infty\frac{\Delta^k}{k!}\Big\{\frac{\partial^k}{\partial\Delta^k}\Gamma(\Delta,\varepsilon)\Big|_{\Delta=0}\Big\}\Big|.\end{aligned}\tag{16}$$

As $\Gamma(\Delta,0)$ is a meromorphic function with a simple pole in 0 and residue 1, there exists a sequence $(a_k)_{k\geq0}$ such that $\Gamma(\Delta)=\frac{1}{\Delta}+\sum_{k=0}^\infty a_k\Delta^k$. Therefore,

$$1-\Delta\Gamma(\Delta,0)=\Delta\sum_{k=0}^\infty a_k\Delta^k,$$

and

$$\frac{1-\Delta\Gamma(\Delta)}{\Delta\Gamma(\Delta)}=\frac{\Delta\sum_{k=0}^\infty a_k\Delta^k}{1+\Delta\sum_{k=0}^\infty a_k\Delta^k}=O(\Delta).$$

Let us now study the term $\sum_{k=1}^\infty\frac{\Delta^k}{k!}\big(\frac{\partial^k}{\partial\Delta^k}\Gamma(\Delta,\varepsilon)\big)\big|_{\Delta=0}$. We have:

$$\begin{aligned}\Big|\frac{\partial^k}{\partial\Delta^k}\Gamma(\Delta,\varepsilon)\Big|_{\Delta=0}\Big|&\leq\Big|e^{-1}\int_\varepsilon^1 x^{-1}(\log(x))^kdx\Big|+\Big|\int_1^\infty e^{-x}(\log(x))^kdx\Big|\\&=e^{-1}\frac{|\log(\varepsilon)|^{k+1}}{k+1}+\int_1^\infty e^{-x}(\log(x))^kdx.\end{aligned}$$

Let x_0 be the largest real number such that $e^{\frac{x_0}{2}}=(\log(x_0))^k$. This equation has two solutions if and only if $k\geq6$. If no such point exists, take $x_0=1$. Then,

$$\begin{aligned}\int_1^\infty e^{-x}(\log(x))^kdx&\leq\int_1^{x_0}e^{-x}(\log(x))^kdx+\int_{x_0}^\infty e^{-\frac{x}{2}}dx\leq(\log(x_0))^k\big(e^{-1}-e^{-x_0}\big)+2e^{-\frac{x_0}{2}}\\&\leq e^{\frac{x_0}{2}-1}+e^{-\frac{x_0}{2}}\leq k^k+1,\end{aligned}$$

where we used the inequality $x_0 < 2k\log k$, for each integer k. Summing up, we get

$$\begin{aligned}\Big|\sum_{k=1}^{\infty}\frac{\Delta^k}{k!}\Big\{\frac{\partial^k}{\partial\Delta}\Gamma(\Delta,\varepsilon)\Big|_{\Delta=0}\Big\}\Big| &\leq e^{-1}\sum_{k=1}^{\infty}\frac{\Delta^k}{k!}\frac{|\log(\varepsilon)|^{k+1}}{k+1}+\sum_{k=1}^{5}2e^{-\frac{1}{2}}\frac{\Delta^k}{k!}+\sum_{k=6}^{\infty}\frac{\Delta^k}{k!}(k^k+1)\\ &\leq |\log(\varepsilon)|\big[e^{\Delta|\log(\varepsilon)|}-1\big]+\sum_{k=6}^{\infty}\frac{\Delta^{\frac{k}{2}}}{k!}\Big(\frac{k}{e}\Big)^k+O(\Delta)\\ &\leq (\log(\varepsilon))^2\Delta+O(\Delta).\end{aligned}$$

In the last two steps, we have used first that $\Delta < e^{-2}$ and then the Stirling approximation formula to deduce that the last remaining sum is $O(\Delta^3)$. Clearly, the factor $\frac{1}{\Delta\Gamma(\Delta)}\sim 1$, as $\Delta\to 0$, in (16) does not change the asymptotic. Finally we derive that

$$\Big|\lambda_\varepsilon-\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}\Big| = O\big(\log(\varepsilon)^2\Delta\big).$$

Another consequence is that there exists a constant C, independent of Δ and ε, such that

$$\mathbb{P}(X_\Delta>\varepsilon)\leq\Delta\big(\lambda_\varepsilon+C\log(\varepsilon)^2\Delta\big).$$

We have just established the following result.

Proposition 2 (Gamma Process)**.** *For all $\varepsilon\in(0,1)$, there exist constants C_1 and C_2, only depending on p, such that, for $\Delta>0$ small enough*

$$\begin{aligned}\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon}-\lambda|^p\big]&\leq C_1\Big\{\Big(\frac{\lambda_\varepsilon+\log(\varepsilon)^2\Delta}{n\Delta}\Big)^{\frac{p}{2}}+(\log(\varepsilon)^2\Delta)^p\Big\},\quad \textit{when } p\in[1,2),\\ \mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon}-\lambda|^p\big]&\leq C_2\Big\{\frac{1}{(n\Delta)^{p-1}}\vee\Big(\frac{\lambda_\varepsilon+\log(\varepsilon)^2\Delta}{n\Delta}\Big)^{\frac{p}{2}}+(\log(\varepsilon)^2\Delta)^p\Big\},\quad \textit{when } p\geq 2.\end{aligned}$$

Cauchy process. Let X be a 1-stable Lévy process with

$$f(x)=\frac{1}{\pi x^2}\mathbb{1}_{\mathbb{R}\setminus\{0\}}\quad\text{and}\quad\mathbb{P}(|X_\Delta|>\varepsilon)=2\int_{\frac{\varepsilon}{\Delta}}^{\infty}\frac{dx}{\pi(x^2+1)}.$$

Then, under the asymptotic $\Delta/\varepsilon\to 0$, we have

$$\Big|\frac{\mathbb{P}(|X_\Delta|>\varepsilon)}{\Delta}-\lambda_\varepsilon\Big| = O\Big(\frac{\Delta^2}{\varepsilon^3}\Big).\tag{17}$$

Indeed, observe that with such a choice of the Lévy density we have $\lambda_\varepsilon=\frac{2}{\pi\varepsilon}$ and, furthermore, $\mathbb{P}(|X_\Delta|>\varepsilon)=\frac{2}{\pi}\big(\frac{\pi}{2}-\arctan\big(\frac{\varepsilon}{\Delta}\big)\big)$. Hence, in order to prove (17), it is enough to show that

$$\lim_{\frac{\Delta}{\varepsilon}\to 0}\frac{2}{\pi}\Big|\frac{\varepsilon^3}{\Delta^3}\Big(\frac{\pi}{2}-\arctan\Big(\frac{\varepsilon}{\Delta}\Big)\Big)-\frac{\varepsilon^2}{\Delta^2}\Big|<\infty.\tag{18}$$

To that purpose, we set $y=\frac{\Delta}{\varepsilon}$ and we compute the limit in (18) by means of de l'Hôpital rule:

$$\frac{2}{\pi}\lim_{y\to 0}\left|\frac{1}{y^3}\Big(\frac{\pi}{2}-\arctan\Big(\frac{1}{y}\Big)\Big)-\frac{1}{y^2}\right|=\frac{2}{\pi}\lim_{y\to 0}\left|\frac{\frac{\pi}{2}-\arctan\big(\frac{1}{y}\big)-y}{y^3}\right|=\lim_{y\to 0}\frac{y^2}{(1+y^2)3\pi y^2}<\infty.$$

Therefore, in the case where X is a Cauchy process of parameters $(1,1)$, Theorem 2 gives:

Proposition 3 (Cauchy Process)**.** *Let $0<\varepsilon=\varepsilon_n\leq 1$ and $\Delta=\Delta_n$ be such that $\lim_{n\to\infty}\frac{\Delta_n}{\varepsilon_n}=0$. Then, for $p\geq 1$ there exist constants C_1, C_2 and n_0, depending only on p, such that $\forall n\geq n_0$*

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon}-\lambda|^p\big]\leq C_1\frac{\Delta^{2p}}{\varepsilon^{3p}}+(n\Delta)^{-\frac{p}{2}}\Big(\frac{1}{\varepsilon}+C_2\frac{\Delta^2}{\varepsilon^3}\Big).$$

Inverse Gaussian process. Let X be an inverse Gaussian process of parameter $(1,1)$, i.e.

$$f(x)=\frac{e^{-x}}{x^{\frac{3}{2}}}\mathbb{1}_{(0,\infty)}(x)\quad\text{ and }\quad\mathbb{P}(X_\Delta>\varepsilon)=\Delta e^{2\Delta\sqrt{\pi}}\int_\varepsilon^\infty\frac{e^{-x-\frac{\pi\Delta^2}{x}}}{x^{\frac{3}{2}}}dx.$$

Then,

$$\left|\frac{\mathbb{P}(X_\Delta>\varepsilon)}{\Delta}-\lambda_\varepsilon\right|\leq\left|e^{2\Delta\sqrt{\pi}}\int_\varepsilon^\infty\frac{e^{-x}\big(e^{-\frac{\pi\Delta^2}{x}}-1\big)}{x^{\frac{3}{2}}}dx\right|+\big(e^{2\Delta\sqrt{\pi}}-1\big)\int_\varepsilon^\infty\frac{e^{-x}}{x^{\frac{3}{2}}}dx=:I+II.$$

After writing the exponential $e^{-\frac{\pi\Delta^2}{x}}$ as an infinite sum, we get $I=O\big(\frac{\Delta^2}{\varepsilon^{\frac{3}{2}}}\big)$ if $\Delta\lambda_\varepsilon\propto\frac{\Delta}{\sqrt{\varepsilon}}\to 0$. Expanding $e^{2\Delta\sqrt{\pi}}$ one finds that, under the same hypothesis, $II=O(\Delta\lambda_\varepsilon)=O\big(\frac{\Delta}{\sqrt{\varepsilon}}\big)$. Theorem 2 leads to the following result.

Proposition 4 (Inverse Gaussian Process)**.** *Let $0<\varepsilon=\varepsilon_n\leq 1$ and $\Delta=\Delta_n$ be such that $\lim_{n\to\infty}\frac{\Delta_n}{\sqrt{\varepsilon_n}}=0$. Then for all $p\geq 1$ there exist constants C_1, C_2 and n_0, depending only on p, such that for all $n\geq n_0$*

$$\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon}-\lambda|^p\big]\leq C_1\Big(\frac{\Delta^p}{\varepsilon^{p/2}}+\frac{\Delta^{2p}}{\varepsilon^{3/2}}\Big)+(n\Delta)^{-\frac{p}{2}}\Big(\frac{1}{\sqrt{\varepsilon}}+C_2\Big(\frac{\Delta^2}{\varepsilon^{3/2}}+\frac{\Delta}{\sqrt{\varepsilon}}\Big)\Big)^{\frac{p}{2}}.$$

3.3 Statistical properties of $\widehat{h}_{n,\varepsilon}$

3.3.1 Construction of $\widehat{h}_{n,\varepsilon}$

We estimate the density h_ε using a linear wavelet density estimator and study its performances uniformly over Besov balls (see Kerkyacharian and Picard [26] or Häddle et al. [23]). We state the result and assumptions in terms of the Lévy density f as it is the quantity of interest.

Preliminary on Besov spaces. Let (Φ, Ψ) be a pair of scaling function and mother wavelet which are compactly supported, of class C^r and generate a regular wavelet basis adapted to the estimation interval $A(\varepsilon)$ (e.g. Daubechie's wavelet). Moreover suppose that $\{\Phi(x-k), k \in \mathbb{Z}\}$ is an orthonormal family of $L_2(\mathbb{R})$. For all $f \in L_{p,\varepsilon}$ we write for $j_0 \in \mathbb{N}$

$$f(x) = \sum_{k\in\Lambda_{j_0}} \alpha_{j_0k}(f)\Phi_{j_0k}(x) + \sum_{j\geq j_0}\sum_{k\in\Lambda_j} \beta_{jk}(f)\Psi_{jk}(x), \quad \forall x \in A(\varepsilon)$$

where $\Phi_{j_0k}(x) = 2^{\frac{j_0}{2}}\Phi(2^{j_0}x-k)$, $\Psi_{jk}(x) = 2^{\frac{j}{2}}\Psi(2^jx-k)$ and the coefficients are

$$\alpha_{j_0k}(f) = \int_{A(\varepsilon)} \Phi_{j_0k}(x)f(x)dx \quad \text{and} \quad \beta_{jk}(f) = \int_{A(\varepsilon)} \Psi_{jk}(x)f(x)dx.$$

As we consider compactly supported wavelets, for every $j \geq j_0$, the set Λ_j incorporates boundary terms that we choose not to distinguish in notation for simplicity. In the sequel we apply this decomposition to h_ε. This is justified because $f_\varepsilon \in L_{p,\varepsilon}$ implies $h_\varepsilon \in L_{p,\varepsilon}$ and the coefficients of its decomposition are $\alpha_{j_0k}(h_\varepsilon) = \alpha_{j_0k}(f)/\lambda_\varepsilon$ and $\beta_{j_0k}(h_\varepsilon) = \beta_{j_0k}(f)/\lambda_\varepsilon$. The latter can be interpreted as the expectations of $\Phi_{j_0k}(U)$ and $\Psi_{jk}(U)$ where U is a random variable with density h_ε with respect to the Lebesgue measure.

We define Besov spaces in terms of wavelet coefficients as follows. For $r > s > 0$, $p \in [1, \infty)$ and $1 \leq q \leq \infty$ a function f belongs to the Besov space $B^s_{p,q}(A(\varepsilon))$ if the norm

$$\|f\|_{B^s_{p,q}(A(\varepsilon))} := \Big(\sum_{k\in\Lambda_{j_0}} |\alpha_{j_0k}(f)|^p\Big)^{\frac{1}{p}} + \Big[\sum_{j\geq j_0}\Big(2^{j(s+1/2-1/p)}\Big(\sum_{k\in\Lambda_j}|\beta_{jk}(f)|^p\Big)^{\frac{1}{p}}\Big)^q\Big]^{\frac{1}{q}} \tag{19}$$

is finite, with the usual modification if $q = \infty$. We consider Lévy densities f with respect to the Lebesgue measure, whose restriction to the interval $A(\varepsilon)$ lies into a Besov ball:

$$\mathscr{F}(s, p, q, \mathfrak{M}_\varepsilon, A(\varepsilon)) = \big\{f \in L_{p,\varepsilon} \ : \ \|f\|_{B^s_{p,q}(A(\varepsilon))} \leq \mathfrak{M}_\varepsilon\big\}. \tag{20}$$

Note that the regularity assumption is imposed on $f_{|A(\varepsilon)}$ viewed as a $L_{p,\varepsilon}$ function. Therefore the dependency in $A(\varepsilon)$ (hence $a(\varepsilon)$) lies in the constant $\mathfrak{M}_\varepsilon$. Also, the parameter p measuring the loss of our estimator is the same as the one measuring the Besov regularity of the function, this is discussed in Section 3.3.2. The following lemma is immediate from the definitions of h_ε and the Besov norm (19).

Lemma 2. *Let f be in $\mathscr{F}(s, p, q, \mathfrak{M}_\varepsilon, A(\varepsilon))$, then $h_\varepsilon = \frac{f_\varepsilon}{\lambda_\varepsilon}$ belongs to $\mathscr{F}\big(s, p, q, \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}, A(\varepsilon)\big)$.*

Construction of $\widehat{h}_{n,\varepsilon}$. Consider $\boldsymbol{D}_{n,\varepsilon}$, the increments larger than ε. We need to estimate the jump density h_ε but we only have access to the indirect observations $\{X_{i\Delta} - X_{(i-1)\Delta}\ i \in \mathscr{I}_\varepsilon\}$, where for each $i \in \mathscr{I}_\varepsilon$, we have

$$X_{i\Delta} - X_{(i-1)\Delta} = M_{i\Delta}(\varepsilon) - M_{(i-1)\Delta}(\varepsilon) + b_\nu(\varepsilon) + Z_{i\Delta}(\varepsilon) - Z_{(i-1)\Delta}(\varepsilon).$$

The problem is twofold. First, there is a deconvolution problem as the information on h_ε is contained in the observations $\{Z_{i\Delta}-Z_{(i-1)\Delta}, i\in\mathscr{I}_\varepsilon\}$. The distribution of the noise $M_\Delta(\varepsilon)+b_\nu(\varepsilon)$ is unknown, but since this quantity is small we neglect this noise and make the approximation:

$$X_{i\Delta}-X_{(i-1)\Delta}\approx Z_{i\Delta}(\varepsilon)-Z_{(i-1)\Delta}(\varepsilon),\quad \forall i\in\mathscr{I}_\varepsilon. \tag{21}$$

Second, even overlooking that it is possible that for some $i_0\in\mathscr{I}_\varepsilon$, $|X_{i_0\Delta}-X_{(i_0-1)\Delta}|>\varepsilon$ and $Z_{i_0\Delta}-Z_{(i_0-1)\Delta}=0$, the common density of $Z_{i\Delta}-Z_{(i-1)\Delta}|Z_{i\Delta}-Z_{(i-1)\Delta}\neq 0$ is not h_ε but it is given by

$$p_{\Delta,\varepsilon}(x)=\sum_{k=1}^\infty \mathbb{P}(N_\Delta(\varepsilon)=k|N_\Delta(\varepsilon)\neq 0)h_\varepsilon^{\star k}(x)=\sum_{k=1}^\infty\frac{(\lambda_\varepsilon\Delta)^k}{k!(e^{\lambda_\varepsilon\Delta}-1)}h_\varepsilon^{\star k}(x),\quad \forall x\in\mathbb{R}, \tag{22}$$

where $\star$ denotes the convolution product. However, in the asymptotic $\Delta\to 0$, we can neglect the possibility that more than one jump of $N(\varepsilon)$ occurred in an interval of length Δ. Indeed, we have the following lemma.

Lemma 3. *If $\lambda_\varepsilon\Delta\to 0$, then for all $p\geq 1$ there exists some constant $C>2$ such that:*

$$\big\|p_{\Delta,\varepsilon}-h_\varepsilon\big\|_{L_{p,\varepsilon}}\leq C\Delta\|f\|_{L_{p,\varepsilon}}.$$

Finally, our estimator is based on the chain of approximations

$$h_\varepsilon\approx p_{\Delta,\varepsilon}\approx\mathcal{L}(X_\Delta||X_\Delta|>\varepsilon).$$

Therefore, we consider the following estimator

$$\widehat{h}_{n,\varepsilon}(x)=\sum_{k\in\Lambda_J}\widehat{\alpha}_{J,k}\Phi_{Jk}(x),\quad x\in A(\varepsilon), \tag{23}$$

where J is an integer to be chosen and

$$\widehat{\alpha}_{J,k}:=\frac{1}{\boldsymbol{n}(\varepsilon)}\sum_{i\in\mathscr{I}_\varepsilon}\Phi_{Jk}(X_{i\Delta}-X_{(i-1)\Delta}).$$

We work with a linear estimator, despite the fact that they are not always minimax for general Besov spaces $B^s_{\pi,q}$, $1\leq\pi,q\leq\infty$ ($\pi\neq p$). Our choice is motivated by the fact that, contrary to adaptive optimal wavelet threshold estimators, linear estimators permit to estimate densities on non-compact intervals. But, most importantly, to evaluate the loss due to the fact that we neglect the small jumps $M_\Delta(\varepsilon)$ (see (21)), we make an approximation at order 1 of our estimator $\widehat{h}_{n,\varepsilon}$. We need our estimator to depend smoothly on the observations, which is not the case if we consider usual thresholding methods. Finally, we recall that on the class $\mathscr{F}(s,p,q,\mathfrak{M}_\varepsilon,A(\varepsilon))$ this estimator is optimal in the context of density estimation from direct i.i.d. observations (see Kerkyacharian and Picard [26], Theorem 3).

3.3.2 Upper bound results

Adapting the results of [26], we derive the following conditional upper bound for the estimation of h_ε when the Lévy measure is infinite. The case where X is a compound Poisson process is illustrated in Proposition 5. Recall that $A(\varepsilon) = (-\overline{A}, -a(\varepsilon)] \cup [a(\varepsilon), \overline{A})$ with $\overline{A} \in [1, \infty]$.

Theorem 3. *Assume that f belongs to the functional class $\mathscr{F}(s,p,q,\mathfrak{M}_\varepsilon, A(\varepsilon))$ defined in (20), for some $1 \le q \le \infty$, $1 \le p < \infty$, $\varepsilon \in (0,1]$ and $\overline{A} < \infty$. Let $r > s > \frac{1}{p}$, $\widehat{h}_{n,\varepsilon}$ be the wavelet estimator of h_ε on $A(\varepsilon)$, defined in (23). Let $v_\Delta(\varepsilon) := \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)| > \varepsilon)$, $F_\Delta(\varepsilon) := \mathbb{P}(|X_\Delta| > \varepsilon)$ and $\sigma^2(\varepsilon) := \int_{|x|\le\varepsilon} x^2 \nu(dx)$. If $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \le \frac{1}{3}$ and $\lambda_\varepsilon \Delta \to 0$ as $n \to \infty$, then the following inequality holds. For all $J \in \mathbb{N}$ and for all finite $p \ge 2$, there exists a positive constant $C > 0$ such that:*

$$\begin{aligned}\mathbb{E}\big(\|\widehat{h}_{n,\varepsilon}(\{X_{i\Delta} - X_{(i-1)\Delta}\}_{i\in\mathscr{I}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p | \mathscr{I}_\varepsilon\big) \le C\bigg\{ & 2^{2Jp}\Big[\Big(\frac{v_\Delta(\varepsilon)}{\boldsymbol{n}(\varepsilon)F_\Delta(\varepsilon)}\Big)^{p/2} + \Big(\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)}\Big)^p\Big] \\ & + \Big[\Big(2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} 2^{Jp/2}\boldsymbol{n}(\varepsilon)^{-p} + (\Delta\|f\|_{L_{p,\varepsilon}})^p\Big] \\ & + 2^{J(5p/2-1)}\Big[\boldsymbol{n}(\varepsilon)^{1-p}\Delta + \boldsymbol{n}(\varepsilon)^{-p/2}\big(\sigma^2(\varepsilon)\Delta\big)^{p/2} + (b_\nu(\varepsilon)\Delta)^p\Big]\bigg\},\end{aligned}$$

where $\boldsymbol{n}(\varepsilon)$ denotes the cardinality of $\mathscr{I}_\varepsilon$ and C only depends on s, p, $\|h_\varepsilon\|_{L_{p,\varepsilon}}$, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. For $1 \le p < 2$ this bound still holds if one requires in addition that $h_\varepsilon(x) \le w(x)$, $\forall x \in \mathbb{R}$ for some symmetric function $w \in L_{p/2}$.

The assumption $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \le \frac{1}{3}$ is not restrictive: this term is required to tend to 0 to get a consistent procedure. An immediate consequence of the proof of Theorem 3 is the following.

Proposition 5. *Assume that f is the Lévy density of a compound Poisson process and that it belongs to the functional class $\mathscr{F}(s,p,q,\mathfrak{M}_0, \mathbb{R}\setminus\{0\})$ defined in (20), for some $1 \le q \le \infty$, $1 \le p < \infty$. Take $\overline{A} = \infty$ and let $r > s > \frac{1}{p}$, $\widehat{h}_{n,0}$ be the wavelet estimator of h_0 on $\mathbb{R}\setminus\{0\}$, defined in (23). Then, for all $J \in \mathbb{N}$ and $p \in [2,\infty)$, there exists $C > 0$ such that:*

$$\mathbb{E}\big(\|\widehat{h}_{n,0}(\{X_{i\Delta}-X_{(i-1)\Delta}\}_{i\in\mathscr{I}_0}) - h_0\|_{L_{p,0}}^p | \mathscr{I}_0\big) \le C\Big[2^{-Jsp} + 2^{Jp/2}\boldsymbol{n}(0)^{-p} + (\Delta\|f\|_{L_{p,0}})^p\Big],$$

where $\boldsymbol{n}(0)$ is the cardinality of $\mathscr{I}_0$ and C depends on s, p, $\|h_0\|_{L_{p,0}}$, $\|h_0\|_{L_{p/2,0}}$, λ_0, $\mathfrak{M}_0$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. For $1 \le p < 2$ this bound still holds if one requires in addition that $h_0(x) \le w(x)$, $\forall x \in \mathbb{R}$ for some symmetric function $w \in L_{p/2}$.

Taking J such that $2^J = \boldsymbol{n}(0)^{\frac{1}{2s+1}}$ leads to an upper bound in $\boldsymbol{n}(0)^{-\frac{s}{2s+1}} \vee \Delta$, where $\boldsymbol{n}(0)^{-\frac{s}{2s+1}}$ is the optimal rate of convergence for the density estimation problem from $\boldsymbol{n}(0)$ i.i.d. direct observations. The error rate Δ is due to the omission of the event that more than one jump may occur in an interval of length Δ.

To get an unconditional bound we introduce the following result.

Lemma 4. *Let $F_\Delta(\varepsilon) := \mathbb{P}(|X_\Delta| > \varepsilon)$. For all $r \geq 0$ we have*

$$\Big(\frac{3nF_\Delta(\varepsilon)}{2}\Big)^{-r} \leq \mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\big) \leq 2\exp\big(\tfrac{-3nF_\Delta(\varepsilon)}{32}\big) + \Big(\frac{nF_\Delta(\varepsilon)}{2}\Big)^{-r}.$$

Using Lemma 4, together with (12) , we can remove the conditioning on $\mathscr{I}_\varepsilon$ and we get an unconditional upper bound for $\widehat{h}_{n,\varepsilon}$.

Corollary 1. *Assume that f belongs to the functional class $\mathscr{F}(s,p,q,\mathfrak{M}_\varepsilon, A(\varepsilon))$ defined in (20), for some $1 \leq q \leq \infty$, $1 \leq p < \infty$ and $\overline{A} < \infty$. Let $r > s > \frac{1}{p}$ and let $\widehat{h}_{n,\varepsilon}$ be the wavelet estimator of h_ε on $A(\varepsilon)$, defined in (23). If $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \leq \frac{1}{3}$, $\lambda_\varepsilon\Delta \to 0$ and $nF_\Delta(\varepsilon) \to \infty$ as $n \to \infty$, then, for all $J \in \mathbb{N}$ and $p \in [2,\infty)$ the following inequality holds:*

$$\begin{aligned}\mathbb{E}\big(\|\widehat{h}_{n,\varepsilon}(\{X_{i\Delta} - X_{(i-1)\Delta}\}_{i\in\mathscr{I}_\varepsilon}) - h_\varepsilon\|^p_{L_{p,\varepsilon}}\big) \leq C\Big\{&2^{2Jp}\Big[\Big(\frac{v_\Delta(\varepsilon)}{nF_\Delta(\varepsilon)^2}\Big)^{p/2} + \Big(\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)}\Big)^p\Big]\\ &+ \Big[\Big(2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}2^{Jp/2}\big(nF_\Delta(\varepsilon)\big)^{-p} + (\Delta\|f\|_{L_{p,\varepsilon}})^p\Big]\\ &+ 2^{J(5p/2-1)}\Big[(nF_\Delta(\varepsilon))^{1-p}\Delta + (nF_\Delta(\varepsilon))^{-p/2}\big(\sigma^2(\varepsilon)\Delta\big)^{p/2} + (b_\nu(\varepsilon)\Delta)^p\Big]\Big\},\end{aligned}$$

for some $C > 0$ depending only on s, p, $\|h_\varepsilon\|_{L_{p,\varepsilon}}$, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. For $1 \leq p < 2$ this bound still holds if one requires in addition that $h_\varepsilon(x) \leq w(x)$, $\forall x \in \mathbb{R}$ for some symmetric function $w \in L_{p/2}$.

The various terms appearing in this upper bound are discussed in Section 4.1. The implied rates are illustrated on examples in Section 4.2.

4 Statistical properties of $\widehat{f}_{n,\varepsilon}$

Combining the results in Theorem 2 and Corollary 1 we derive the following upper bound for the estimator $\widehat{f}_{n,\varepsilon}$ of the Lévy density f when $\nu(\mathbb{R}) = \infty$. The case where X is a compound Poisson process is illustrated in Proposition 6.

Theorem 4. *Let f belong to the functional class $\mathscr{F}(s,p,q,\mathfrak{M}_\varepsilon, A(\varepsilon))$ defined in (20), for some $1 \leq q \leq \infty$, $1 \leq p < \infty$, $\varepsilon \in (0,1]$ and $\overline{A} < \infty$. Let $r > s > \frac{1}{p}$ and let $\widehat{f}_{n,\varepsilon}$ be the estimator of f on $A(\varepsilon)$, defined in (10). Then, under the assumptions $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \leq \frac{1}{3}$, $\lambda_\varepsilon\Delta \to 0$ and $nF_\Delta(\varepsilon) \to \infty$ as $n \to \infty$, for all $J \in \mathbb{N}$ and $p \in [2,\infty)$, there exists $C > 0$ such that the following inequality holds:*

$$\begin{aligned}\big[\ell_{p,\varepsilon}(\widehat{f}_{n,\varepsilon}, f)\big]^p \leq C\Big\{&\Big[\Big(\frac{F_\Delta(\varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}} + \Big|\lambda_\varepsilon - \frac{F_\Delta(\varepsilon)}{\Delta}\Big|^p\Big]\Big(\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p\\ &+ \lambda_\varepsilon^p\Big\{2^{2Jp}\Big[\Big(\frac{v_\Delta(\varepsilon)}{nF_\Delta(\varepsilon)^2}\Big)^{p/2} + \Big(\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)}\Big)^p\Big]\\ &+ \Big[\Big(2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}2^{Jp/2}\big(nF_\Delta(\varepsilon)\big)^{-p} + (\Delta\|f\|_{L_{p,\varepsilon}})^p\Big]\\ &+ 2^{J(5p/2-1)}\Big[(nF_\Delta(\varepsilon))^{1-p}\Delta + (nF_\Delta(\varepsilon))^{-p/2}\big(\sigma^2(\varepsilon)\Delta\big)^{p/2} + (b_\nu(\varepsilon)\Delta)^p\Big]\Big\}\Big\}\end{aligned}$$

where $v_\Delta(\varepsilon) := \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)| > \varepsilon)$, $F_\Delta(\varepsilon) := \mathbb{P}(|X_\Delta| > \varepsilon)$, $\sigma^2(\varepsilon) := \int_{|x|\leq\varepsilon} x^2\nu(dx)$ and C depends on s, p, $\|h_\varepsilon\|_{L_{p,\varepsilon}}$, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. For $1 \leq p < 2$ this bound still holds if one requires in addition that $f_\varepsilon(x) \leq w(x)$, $\forall x \in \mathbb{R}$ for some symmetric function $w \in L_{p/2}$.

4.1 Discussion

The upper bound presented in Theorem 4 is difficult to interpret in general. Here, we give a rough intuition of what terms are dominating and where they come from. Thinking back on our strategy, we made different approximations that entail four different sources of errors (points 2-3-4 are related to the estimation of h_ε whereas point 1 to the estimation of λ_ε).

1. **Estimation of λ_ε:** In Section 3.2.2 we have already discussed this point. Our approximation strategy for the intensity λ_ε leads to the error
$$\Big(\frac{F_\Delta(\varepsilon)}{n\Delta^2}\Big)^{\frac{p}{2}} + \Big|\lambda_\varepsilon - \frac{F_\Delta(\varepsilon)}{\Delta}\Big|^p := E_1.$$

2. **Neglecting the event $\{|M_\Delta(\varepsilon)+b_\nu(\varepsilon)| > \varepsilon\}$:** We consider that each time an increment X_Δ exceeds the threshold ε the associated Poisson process $N_\Delta(\varepsilon)$ is nonzero. This leads to the error
$$2^{2J}\Bigg\{\sqrt{\frac{v_\Delta(\varepsilon)}{nF_\Delta(\varepsilon)^2}} + \frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)}\Bigg\} \asymp 2^{2J}\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} := E_2.$$
This error is unavoidable as we do not observe $M(\varepsilon)$ and $Z(\varepsilon)$ separately.

3. **Neglecting the presence of $M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)$:** In (21) we ignore the convolution structure of the observations. This produces the error in
$$2^{J(5/2-1/p)}\Big\{(nF_\Delta(\varepsilon))^{-1+1/p}\Delta + (nF_\Delta(\varepsilon))^{-1/2}\big(\sigma^2(\varepsilon)\Delta\big)^{1/2} + (b_\nu(\varepsilon)\Delta)^p\Big\} := E_3.$$
It would have been difficult to have a better strategy than neglecting $M_\Delta(\varepsilon) + b_\nu(\varepsilon)\Delta$: the distribution of $M_\Delta(\varepsilon)$ is unknown, then we cannot take into account the convolution structure of the observations. Moreover, even if we did know it (or could estimate it), deconvolution methods are essentially adapted to L_2 losses.

4. **Estimation of the compound Poisson $Z(\varepsilon)$:** This estimation problem is solved in two steps. First, we neglect the event $\{N_\Delta(\varepsilon) \geq 2\}$ which generates the error:
$$\Delta\|f\|_{L_{p,\varepsilon}} := E_4.$$
This error could have been improved considering a corrected estimator as in [14], but this would have added even more heaviness in the final result. Second, we recover an estimation error that is classical for the density estimation problem from i.i.d. observations in
$$2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}2^{J/2}\big(nF_\Delta(\varepsilon)\big)^{-1} := E_5.$$

One can get easily convinced that the most significant term is E_2. Using (12) we see that it is possible to choose J, going to infinity, such that E_2 still tends to 0. This choice of J together with $\Delta \to 0$ and $nF_\Delta(\varepsilon) \to 0$ leads to an upper bound that goes to 0. Balancing these five terms to get an explicit rate is difficult without further assumptions. But, in general, the leading term will be imposed by the unknown rate of convergence of $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)}$ to 0 of and consequently by E_2. We cannot ensure that this rate is optimal and we cannot propose an adaptive choice of J as, in practice, as we already underlined, a sharp control of $v_\Delta(\varepsilon)$ is not known. Below we discuss the main problems related with this quantity $v_\Delta(\varepsilon)$.

4.1.1 How to control the small jumps of a Lévy process

As we have already pointed out, a crucial role in determining the rate of convergence of our estimators is played by the quantity $v_\Delta(\varepsilon) := \mathbb{P}(|M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon)| > \varepsilon)$. In the literature papers devoted to expansions for the distributions of Lévy processes already exists (see, e.g., [33] and [18]) but they cannot be used in our framework. The expansions for $\mathbb{P}(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon) > x)$ holds only for x large enough with respect to ε.

A theoretical approach to compute $v_\Delta(\varepsilon)$ is offered by the inversion formula and the Lévy-Khintchine formula. We reproduce the computations only in the case where X is a subordinator but they can be done in general. Formally, let X be a subordinator with Lévy measure ν, we have

$$M_t(\varepsilon) + tb_\nu(\varepsilon) = \sum_{s\le t} \Delta X_s \mathbb{I}_{0\le \Delta X_s \le \varepsilon}.$$

By the Lévy-Khintchine formula, it follows that

$$\mathbb{E}\Big[e^{iu(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon))}\Big] = \exp\bigg(\Delta\int_0^\varepsilon (e^{iuy}-1)\nu(dy)\bigg) := \varphi(u)$$

and we can express the density of the random variable $M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon)$ as

$$d(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \exp\bigg(-iux + \Delta\int_0^\varepsilon (e^{iuy}-1)\nu(dy)\bigg)du.$$

Therefore

$$\mathbb{P}(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon) > \varepsilon) = \frac{1}{2\pi}\int_\varepsilon^\infty\int_{\mathbb{R}} \exp\bigg(-iux + \Delta\int_0^\varepsilon (e^{iuy}-1)\nu(dy)\bigg)dudx.$$

Unfortunately, the double integral above is far from being easily computable. Another possible representation that one could use is the one provided in [19]:

$$\begin{aligned}\mathbb{P}(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon) > \varepsilon) &= \frac{1}{2} - \frac{1}{2\pi}\int_0^\infty \frac{e^{-it\varepsilon}\varphi(-t) - e^{it\varepsilon}\varphi(t)}{it}dt\\ &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{\mathrm{Im}\big[e^{-iu\varepsilon}\varphi(u)\big]}{u}du\\ &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \frac{e^{\Delta\int_0^\varepsilon(\cos(uy)-1)\nu(dy)}\sin\Big(\Delta\int_0^\varepsilon \sin(uy)\nu(dy) - u\varepsilon\Big)}{u}du,\end{aligned}$$

but, again, these expressions are hard to handle in practice.

However, at least in the case where X is a subordinator, something more precise can be said about $v_\Delta(\varepsilon)$ thanks to the relation (13)

$$\mathbb{P}(M_\Delta(\varepsilon)+\Delta b_\nu(\varepsilon)>\varepsilon)=e^{\lambda_\varepsilon\Delta}\big[\mathbb{P}(X_\Delta>\varepsilon)+e^{-\lambda_\varepsilon\Delta}-1\big],\quad \varepsilon>0.$$

In particular, for the class of Gamma processes and Inverse Gaussian processes treated in Section 3.2.3, we have

$$v_\Delta(\varepsilon)=e^{\lambda_\varepsilon\Delta}\big[(\mathbb{P}(X_\Delta>\varepsilon)-\lambda_\varepsilon\Delta)+(e^{-\lambda_\varepsilon\Delta}-1+\lambda_\varepsilon\Delta)\big]=O(\lambda_\varepsilon^2\Delta^2)$$

as $\lambda_\varepsilon\Delta\to 0$.

4.2 Examples

We go back to the first two examples developed in Section 3.2.3.

Compound Poisson process (*Continued*). In this case we take $\varepsilon=0$ as $\lambda:=\lambda_0<\infty$ and we have

$$F_\Delta(0)=\mathbb{P}(|X_\Delta|>0)=1-e^{-\lambda\Delta}=O(\lambda\Delta).$$

It is straightforward to see that $nF_\Delta(0)\to\infty$. Moreover, the choice $\varepsilon=0$, simplifies the proof of Theorem 3 significantly. Indeed, we have that $v_\Delta(0)=0$, $\mathscr{I}_0=\mathscr{K}_0$, $\boldsymbol{n}(0)=\widetilde{\boldsymbol{n}}(0)$ and that X_Δ has distribution $p_{\Delta,0}$ (see Section 5.3). Proposition 5 and Lemma 4 lead then to the following upper bound. For all $J\in\mathbb{N}$, $\forall h_0\in\mathscr{F}\big(s,p,q,\frac{\mathfrak{M}_0}{\lambda},\mathbb{R}\setminus\{0\}\big)$, $1\le q\le\infty$ and $p\ge 1$, there exists a constant $C>0$ such that:

$$\mathbb{E}\big(\|\widehat{h}_{n,0}-h_0\|_{L_{p,0}}^p\big)\le C\big\{2^{-Jsp}+2^{Jp/2}(n\Delta)^{-p/2}+\Delta^p\big\}$$

where C depends on λ, $\mathfrak{M}_0$, s, p, $\|h_0\|_{L_{p,0}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. Choosing J such that $2^J=(n\Delta)^{\frac{1}{2s+1}}$ we get

$$\mathbb{E}\big(\|\widehat{h}_{n,0}-h_0\|_{L_{p,0}}^p\big)\le C\big\{(n\Delta)^{-\frac{sp}{2s+1}}+\Delta^p\big\},$$

where the first term is the optimal rate of convergence to estimate $p_{\Delta,0}$ from the observations $\boldsymbol{D}_{n,0}$ and the second term is the deterministic error of the approximation of h_0 by $p_{\Delta,0}$. This result is consistent with the results in [14] and is more general in the sense that the estimation interval is unbounded.

Concerning the estimation of the Lévy density $f=f_0$, we apply a slight modification of Theorem 4 (due to the simplifications that occur when taking $\varepsilon=0$), and we use Proposition 1 to derive the following result. Let $\varepsilon=0$, assume that f_0 belongs to the class $\mathscr{F}(s,p,q,\mathfrak{M}_0,[-\overline{A},\overline{A}]\setminus\{0\})$ defined in (20), where $1\le q\le\infty$, $1\le p<\infty$ and $\overline{A}<\infty$. Here we consider a bounded set $A(0)$ for technical reasons (see the proof of Theorem 4), this assumptions might be removed at the expense of additional technicalities. Let J be such that $2^J=(n\Delta)^{\frac{1}{2s+1}}$, then for $p\ge 2$ we have

$$\big[\ell_{p,0}\big(\widehat{f}_{n,0},f\big)\big]^p=O\Big((n\Delta)^{1-p}\vee(n\Delta)^{-\frac{p}{2}}+(n\Delta)^{-\frac{sp}{2s+1}}+\Delta^p\Big)=O\big((n\Delta)^{-\frac{sp}{2s+1}}+\Delta^p\big)$$

The case $p \in [1,2)$ can be treated similarly and leads to the same rate. As earlier, the first term is the optimal rate of convergence to estimate $p_{\Delta,0}$ from the observations $\boldsymbol{D}_{n,0}$ and the second term gathers the deterministic errors of the approximations of h_0 by $p_{\Delta,0}$ and λ_0 by $\frac{1}{\Delta}\mathbb{P}(|X_\Delta| > 0)$. We therefore established the following result.

Proposition 6. *Let $f \in \mathscr{F}(s,p,q,\mathfrak{M}_0,[-\overline{A},\overline{A}]\setminus\{0\})$, $1 \le q \le \infty$, $1 \le p < \infty$ and $\overline{A} < \infty$, be the Lévy density of a compound Poisson process. Let $r > s > \frac{1}{p}$ and let $\widehat{f}_{n,0}$ be the estimator of $f = f_0$ on $[-\overline{A},\overline{A}]\setminus\{0\}$, defined in (10). Then, for all $p \in [1,\infty)$ and n big enough, there exists a constant $C > 0$ such that*

$$\big[\ell_{p,0}\big(\widehat{f}_{n,0}, f\big)\big]^p \le C\Big((n\Delta)^{-\frac{sp}{2s+1}} + \Delta^p\Big),$$

where C depends on λ, $\mathfrak{M}_0$, s, p, $\|h_0\|_{L_{p,0}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$.

Gamma process (*Continued*). Let X be a Gamma process of parameter $(1,1)$. Let $\varepsilon \in (0,1)$, we have $\lambda_\varepsilon = \int_\varepsilon^\infty \frac{e^{-x}}{x}dx = O(\log(\varepsilon^{-1}))$ and if $\log(\varepsilon)\Delta \to 0$ the above computations lead to

$$F_\Delta(\varepsilon) = \mathbb{P}(X_\Delta > \varepsilon) = O(\lambda_\varepsilon\Delta) \quad \text{and} \quad v_\Delta(\varepsilon) = O(\lambda_\varepsilon^2\Delta^2).$$

Moreover $b_\nu(\varepsilon) = \int_0^\varepsilon x\nu(dx) = O(\varepsilon)$ and $\sigma^2(\varepsilon) = \int_0^\varepsilon x^2\nu(dx) = O(\varepsilon^2)$. Also, we observe that for all $1 \le q \le \infty$ and $1 \le p < \infty$, f_ε belongs to the class $\mathscr{F}(s,p,q,\mathfrak{M}\log(\varepsilon^{-1}),A(\varepsilon))$ for some constant $\mathfrak{M}$. Let $r > s > \frac{1}{p}$, applying Theorem 4 for $p \ge 2$, we derive

$$\big[\ell_{p,\varepsilon}\big(\widehat{f}_{n,\varepsilon}, f\big)\big]^p = O\bigg((\log(\varepsilon^{-1}))^p 2^{-Jsp} + 2^{J(5p/2-1)}\Big[\frac{\Delta\log(\varepsilon^{-1})}{(n\Delta)^{p-1}} + \big(\log(\varepsilon^{-1})\varepsilon\Delta\big)^p\Big]\bigg).$$

Neglecting the effect of ε $\big($e.g. consider $\varepsilon = \frac{1}{\log(n)}\big)$ and setting

$$J = \frac{1}{(sp + \frac{5p}{2} - 1)}\log_2\Big(\frac{\Delta}{(n\Delta)^{p-1}} + \Delta^p\Big),$$

we obtain the following rate of convergence

$$\big[\ell_{p,\varepsilon}\big(\widehat{f}_{n,\varepsilon}, f\big)\big]^p = O\bigg(\Big(\frac{\Delta}{(n\Delta)^{p-1}} + \Delta^p\Big)^{-\frac{s}{(s+5/2-1/p)}}\bigg).$$

If $A(\varepsilon)$ is bounded away from zero (e.g. $a(\varepsilon)$ is non-vanishing), the Lévy density is regular and s can be chosen as large as desired. For large s, we recover the rate $\Delta^p \vee \frac{\Delta}{(n\Delta)^{p-1}}$, which is optimal under the classical condition $\Delta = O(1/\sqrt{n})$.

4.2.1 Conclusion

The accuracy of the estimation of λ_ε and h_ε have already been discussed. Theorem 4 is the aggregate of both results: the rate of $\widehat{f}_{n,\varepsilon}$ is the worst between the two errors. We cannot give a general rate of convergence due to the influence of the small jumps, present in the upper bound via the quantity $v_\Delta(\varepsilon)$ that is difficult to handle in practice. However, consistency of

$\widehat{f}_{n,\varepsilon}$ is ensured for L_p loss functions. Moreover, our upper bounds show clearly the influence of the small jumps. Finally, in the case where X is a compound Poisson process or a Gamma process we recover classical results, which gives credit to the procedure. But the question of whether our procedure is optimal in general, as well as the question of the adaptive choice for J, remain open. Answering them will require a deeper understanding of the quantity $\mathbb{P}(|M_\Delta(\varepsilon)+b_\nu(\varepsilon)|>\varepsilon)$ as $\varepsilon\to 0$.

5 Proofs

In the sequel, C denotes a generic constant whose value may vary from line to line. Its dependencies may be given in indices. The proofs of auxiliary lemmas are postponed to the Appendix in Section 6.

5.1 Proof of Theorem 1

Let $F_\Delta(\varepsilon) := \mathbb{P}(|X_\Delta|>\varepsilon)$ and $\widehat{F}_\Delta(\varepsilon) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta}-X_{(i-1)\Delta}|)$. The following holds

$$\mathbb{E}_{P_n}\Big[\big|\lambda_\varepsilon-\widehat{\lambda}_{n,\varepsilon}\big|^p\Big] \le 2^p\bigg\{\Big|\lambda_\varepsilon-\frac{F_\Delta(\varepsilon)}{\Delta}\Big|^p+\frac{1}{\Delta^p}\mathbb{E}_{P_n}\Big[\big|F_\Delta(\varepsilon)-\widehat{F}_\Delta(\varepsilon)\big|^p\Big]\bigg\}. \tag{24}$$

To control the second term in (24), we introduce the i.i.d. centered random variables

$$U_i := \frac{\mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta}-X_{(i-1)\Delta}|)-F_\Delta(\varepsilon)}{n}, \quad i=1,\dots,n.$$

For $p\ge 2$, an application of the Rosenthal inequality together with $\mathbb{E}\big[|U_i|^p\big]=O\big(\frac{F_\Delta(\varepsilon)}{n^p}\big)$ ensure the existence of a constant C_p such that

$$\mathbb{E}_{P_n}\Big[\Big|\sum_{i=1}^n U_i\Big|^p\Big]\le C_p\Big(n^{1-p}F_\Delta(\varepsilon)+\Big(\frac{F_\Delta(\varepsilon)}{n}\Big)^{p/2}\Big).$$

For $p\in[1,2)$, the Jensen inequality and the previous result for $p=2$ lead to

$$\mathbb{E}_{P_n}\Big[\Big|\sum_{i=1}^n U_i\Big|^p\Big]\le\Big(\mathbb{E}_{P_n}\Big[\Big|\sum_{i=1}^n U_i\Big|^2\Big]\Big)^{p/2}\le\Big(\frac{F_\Delta(\varepsilon)}{n}\Big)^{p/2}.$$

□

5.2 Proof of Theorem 2

Thanks to (24) and using the notations introduced in the proof of Theorem 1, we are only left to show that, for $p\ge 2$,

$$\frac{1}{\Delta^p}\mathbb{E}_{P_n}\Big[\Big|\sum_{i=1}^n U_i\Big|^p\Big]=O\Big(\Big(\frac{F_\Delta(\varepsilon)}{n\Delta^2}\Big)^{p/2}\Big). \tag{25}$$

An application of the Bernstein inequality (using that $|U_i| \leq n^{-1}$ and the fact that the variance $\mathbb{V}[U_i] \leq \frac{F_\Delta(\varepsilon)}{n^2}$) allows us to deduce that

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n U_i\Big| \geq t\Big) \leq 2\exp\Big(-\frac{t^2 n}{2F_\Delta(\varepsilon)+\frac{2t}{3}}\Big).$$

Therefore,

$$\mathbb{E}_{P_n}\Big[\Big|\sum_{i=1}^n U_i\Big|^p\Big] = p\int_0^\infty t^{p-1}\mathbb{P}\Big(\Big|\sum_{i=1}^n U_i\Big| \geq t\Big)dt \leq 2p\int_0^\infty t^{p-1}\exp\Big(-\frac{t^2 n}{2F_\Delta(\varepsilon)+\frac{2t}{3}}\Big)dt.$$

Observe that, for $t \leq \frac{3}{2}F_\Delta(\varepsilon)$, the denominator $2F_\Delta(\varepsilon)+\frac{2t}{3}$ is smaller than $3F_\Delta(\varepsilon)$ while for $t \geq \frac{3}{2}F_\Delta(\varepsilon)$ we have $2F_\Delta(\varepsilon)+\frac{2t}{3} \leq 2t$. It follows that

$$\int_0^\infty t^{p-1}\exp\Big(-\frac{t^2 n}{2F_\Delta(\varepsilon)+\frac{2t}{3}}\Big)dt \leq \int_0^{\frac{3}{2}F_\Delta(\varepsilon)} t^{p-1}\exp\Big(-\frac{t^2 n}{3F_\Delta(\varepsilon)}\Big)dt + \int_{\frac{3}{2}F_\Delta(\varepsilon)}^\infty t^{p-1}e^{-\frac{tn}{2}}dt.$$

After a change of variables, the following inequalities hold:

$$\int_0^{\frac{3}{2}F_\Delta(\varepsilon)} t^{p-1}\exp\Big(-\frac{t^2 n}{3F_\Delta(\varepsilon)}\Big)dt \leq \frac{1}{2}\Big(\frac{3F_\Delta(\varepsilon)}{n}\Big)^{p/2}\Gamma\Big(\frac{p}{2}\Big) \tag{26}$$

and

$$\int_{\frac{3}{2}F_\Delta(\varepsilon)}^\infty t^{p-1}e^{-\frac{tn}{2}}dt \leq \Big(\frac{2}{n}\Big)^p\Gamma\Big(p, \frac{nF_\Delta(\varepsilon)}{4}\Big). \tag{27}$$

Here, $\Gamma(s,x) = \int_x^\infty x^{s-1}e^{-x}dx$ denotes the incomplete Gamma function and $\Gamma(s) = \Gamma(s,0)$ is the usual Gamma function. Equation (26) readily gives the desired asymptotic. To conclude, we use the classical estimate for the incomplete Gamma function for $|x| \to \infty$:

$$\Gamma(s,x) \approx x^{s-1}e^{-x}\Big(1+\frac{s-1}{x}+O\Big(\frac{1}{x^2}\Big)\Big).$$

In particular, when (27) is divided by Δ^p, it is asymptotically $O\big(\frac{1}{(n\Delta)^p}e^{-nF_\Delta(\varepsilon)}\big)$, which goes to 0 faster than (25).

5.3 Proof of Theorem 3

Preliminary. Since the proof of Theorem 3 is lengthy, to help the reader we enlighten here the two main difficulties that arise due to the fact that the estimator $\widehat{h}_{n,\varepsilon}$ uses the observations $\boldsymbol{D}_{n,\varepsilon}$, i.e. $\widehat{h}_{n,\varepsilon} = \widehat{h}_{n,\varepsilon}(\boldsymbol{D}_{n,\varepsilon})$.

1. The cardinality of $\boldsymbol{D}_{n,\varepsilon}$ is $\boldsymbol{n}(\varepsilon)$ that is random. That is why in Theorem 3 we study the risk of this estimator conditionally on $\mathscr{I}_\varepsilon$. We then get the general result using that

   $$\mathbb{E}_{P_n}\Big[\ell_{p,\varepsilon}\big(\widehat{h}_{n,\varepsilon}, h_\varepsilon\big)\Big] = \mathbb{E}\Big[\mathbb{E}_{P_{n,\varepsilon}}\Big[\ell_{p,\varepsilon}\big(\widehat{h}_{n,\varepsilon}, h_\varepsilon\big)\big|\mathscr{I}_\varepsilon\Big]\Big].$$

   Once the conditional expectation is bounded, we use Lemma 4 to remove the conditioning and derive Corollary 1.

2. An observation of $\boldsymbol{D}_{n,\varepsilon}$ is not a realization of h_ε. Indeed, an increment of the process $Z(\varepsilon)$ does not necessarily correspond to one jump, whose density is h_ε, and, more demandingly, the presence of the small jumps $M(\varepsilon)$ needs to be taken into account. To do so we split the sample $\boldsymbol{D}_{n,\varepsilon}$ in two according to the presence or absence of jumps in the Poisson part. On the subsample where the Poisson part is nonzero, we make an expansion at order 1 and we neglect the presence of the small jumps. This is the subject of the following paragraph.

Expansion of $\widehat{h}_{n,\varepsilon}$. Consider $\boldsymbol{D}_{n,\varepsilon} = \{X_{i\Delta} - X_{(i-1)\Delta},\ i \in \mathscr{I}_\varepsilon\}$ the increments larger than ε. Recall that, for each i, we have

$$X_{i\Delta} - X_{(i-1)\Delta} = \Delta b_\nu(\varepsilon) + M_{i\Delta}(\varepsilon) - M_{(i-1)\Delta}(\varepsilon) + Z_{i\Delta}(\varepsilon) - Z_{(i-1)\Delta}(\varepsilon).$$

We split the sample as follows:

$$\mathscr{K}_\varepsilon := \{i \in \mathscr{I}_\varepsilon,\ Z_{i\Delta}(\varepsilon) - Z_{(i-1)\Delta}(\varepsilon) \neq 0\}$$
$$\mathscr{K}_\varepsilon^c := \mathscr{I}_\varepsilon \setminus \mathscr{K}_\varepsilon.$$

Denote by $\widetilde{\boldsymbol{n}}(\varepsilon)$ the cardinality of $\mathscr{K}_\varepsilon$. To avoid cumbersomeness, in the remainder of the proof we write M instead of $M(\varepsilon)$ and Z instead of $Z(\varepsilon)$. Recall that $\Phi_{Jk}(x) = 2^{\frac{J}{2}}\Phi(2^J x - k)$. Using that Φ is continuously differentiable we can write, $\forall k \in \Lambda_J$,

$$\begin{aligned}
\widehat{\alpha}_{J,k} &= \frac{1}{\boldsymbol{n}(\varepsilon)}\Big(\sum_{i\in\mathscr{K}_\varepsilon} + \sum_{i\in\mathscr{K}_\varepsilon^c}\Big)\Phi_{Jk}(X_{i\Delta} - X_{(i-1)\Delta}) \\
&= \frac{1}{\boldsymbol{n}(\varepsilon)}\sum_{i\in\mathscr{K}_\varepsilon}\big\{\Phi_{Jk}(Z_{i\Delta} - Z_{(i-1)\Delta}) + 2^{3J/2}(M_{i\Delta} - M_{(i-1)\Delta} + b_\nu(\varepsilon)\Delta)\Phi'(2^J\eta_i - k)\big\} \\
&\quad + \frac{1}{\boldsymbol{n}(\varepsilon)}\sum_{i\in\mathscr{K}_\varepsilon^c}\Phi_{Jk}(X_{i\Delta} - X_{(i-1)\Delta}),
\end{aligned}$$

where $\eta_i \in [\min\{Z_{i\Delta} - Z_{(i-1)\Delta}, X_{i\Delta} - X_{(i-1)\Delta}\}, \max\{Z_{i\Delta} - Z_{(i-1)\Delta}, X_{i\Delta} - X_{(i-1)\Delta}\}]$. It follows that

$$\begin{aligned}
\widehat{h}_{n,\varepsilon}(x, \{X_{i\Delta} - X_{(i-1)\Delta}\}_{i\in\mathscr{I}_\varepsilon}) &= \sum_{k\in\Lambda_J}\widehat{\alpha}_{J,k}\Phi_{Jk}(x) \\
&:= \frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)}\widetilde{h}_{n,\varepsilon}(x, \{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}) \\
&\quad + \frac{2^{3J/2}}{\boldsymbol{n}(\varepsilon)}\sum_{i\in\mathscr{K}_\varepsilon}(M_{i\Delta} - M_{(i-1)\Delta} + b_\nu(\varepsilon)\Delta)\sum_{k\in\Lambda_J}\Phi'(2^J\eta_i - k)\Phi_{Jk}(x) \\
&\quad + \frac{1}{\boldsymbol{n}(\varepsilon)}\sum_{i\in\mathscr{K}_\varepsilon^c}\sum_{k\in\Lambda_J}\Phi_{Jk}(M_{i\Delta} - M_{(i-1)\Delta} + b_\nu(\varepsilon)\Delta)\Phi_{Jk}(x),
\end{aligned}$$

where conditional on $\mathscr{K}_\varepsilon$, $\widetilde{h}_{n,\varepsilon}(\{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon})$ is the linear wavelet estimator of $p_{\Delta,\varepsilon}$ defined in (22) from $\widetilde{\boldsymbol{n}}(\varepsilon)$ direct measurements. Explicitly, it is defined as follows

$$\widetilde{h}_{n,\varepsilon}(x, \{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}) = \sum_{k\in\Lambda_J}\widetilde{\alpha}_{J,k}\Phi_{Jk}(x), \tag{28}$$

where $\widetilde{\alpha}_{J,k} = \frac{1}{\widetilde{\boldsymbol{n}}(\varepsilon)} \sum_{i\in\mathscr{K}_\varepsilon} \Phi_{Jk}(Z_{i\Delta} - Z_{(i-1)\Delta})$. This is not an estimator as both $\mathscr{K}_\varepsilon$ and $\{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}$ are not observed. However, $\widetilde{\alpha}_{J,k}$ approximates the quantity

$$\alpha_{J,k} := \int_{A(\varepsilon)} \Phi_{Jk}(x) p_{\Delta,\varepsilon}(x) dx. \tag{29}$$

Decomposition of the $L_{p,\varepsilon}$ loss. Taking the $L_{p,\varepsilon}$ norm and applying the triangle inequality we get

$$\begin{aligned}
\|\widehat{h}_{n,\varepsilon}(\{X_{i\Delta} - X_{(i-1)\Delta}\}_{i\in\mathscr{I}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p &\le C_p\bigg\{ \Big(\frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)}\Big)^p \|\widetilde{h}_{n,\varepsilon}(\{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p \\
&\quad + \Big(1 - \frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)}\Big)^p \|h_\varepsilon\|_{L_{p,\varepsilon}}^p \\
&\quad + \frac{2^{3Jp/2}}{\boldsymbol{n}(\varepsilon)^p} \int_{A(\varepsilon)} \Big| \sum_{i\in\mathscr{K}_\varepsilon} (M_{i\Delta} - M_{(i-1)\Delta} + b_\nu(\varepsilon)\Delta) \sum_{k\in\Lambda_J} \Phi'(2^J\eta_i - k)\Phi_{Jk}(x)\Big|^p dx \\
&\quad + \frac{1}{\boldsymbol{n}(\varepsilon)^p} \int_{A(\varepsilon)} \Big| \sum_{i\in\mathscr{K}_\varepsilon^c} \sum_{k\in\Lambda_J} \Phi_{Jk}(M_{i\Delta} - M_{(i-1)\Delta} + b_\nu(\varepsilon))\Phi_{Jk}(x)\Big|^p dx \bigg\} \\
&= C_p\big\{T_1 + T_2 + T_3 + T_4\big\}.
\end{aligned} \tag{30}$$

After taking expectation conditionally on $\mathscr{I}_\varepsilon$ and $\mathscr{K}_\varepsilon$, we bound each term separately.

Remark 1. *If X is a compound Poisson process and we take $\varepsilon = 0$, then $\widehat{h}_{n,\varepsilon} = \widetilde{h}_{n,\varepsilon}$ (and $\boldsymbol{n}(0) = \widetilde{\boldsymbol{n}}(0)$) and $T_2 = T_3 = T_4 = 0$.*

Control of T_1. We have

$$\begin{aligned}
\|\widetilde{h}_{n,\varepsilon}(\{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p &\le C_p\big\{\|\widetilde{h}_{n,\varepsilon}(\{Z_{i\Delta} - Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon}) - p_{\Delta,\varepsilon}\|_{L_{p,\varepsilon}}^p \\
&\quad + \|p_{\Delta,\varepsilon} - h_\varepsilon\|_{L_{p,\varepsilon}}^p\big\} \\
&=: C_p(T_5 + T_6).
\end{aligned}$$

The deterministic term T_6 is bounded using Lemma 3 by $(\Delta\|f\|_{L_{p,\varepsilon}})^p$. Taking expectation conditionally on $\mathscr{I}_\varepsilon$ and $\mathscr{K}_\varepsilon$ of T_5, we recover the linear wavelet estimator of $p_{\Delta,\varepsilon}$ studied by Kerkyacharian and Picard [26] (see their Theorem 2). For the sake of completeness we reproduce the main steps of their proof. First, the control of the bias is the same as in [26], noticing that Lemma 2 implies $p_{\Delta,\varepsilon} \in \mathscr{F}\big(s, p, q, \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}, A(\varepsilon)\big)$ (see Lemma 5.1 in [14]) we get

$$\mathbb{E}\big(T_5|\mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon\big) \le C_p\bigg\{2^{-Jsp}\Big(\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p + 2^{J(p/2-1)} \sum_{k\in\Lambda_J} \mathbb{E}(|\widetilde{\alpha}_{J,k} - \alpha_{J,k}|^p|\mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon)\bigg\},$$

where $\widetilde{\alpha}_{J,k}$ and $\alpha_{J,k}$ are defined in (28) and (29). First consider the case $p \geq 2$. We start by observing that

$$\mathbb{E}\Big[\Big|\frac{1}{\widetilde{\boldsymbol{n}}(\varepsilon)}\sum_{i\in\mathscr{K}_\varepsilon}\Phi_{Jk}(Z_{i\Delta}-Z_{(i-1)\Delta})-\int_{A(\varepsilon)}\Phi_{Jk}(x)p_{\Delta,\varepsilon}(x)\Big|\Big|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big]=$$
$$\sum_{I\subset\{1,\dots,n\}}\mathbb{1}_{\{I=\mathscr{K}_\varepsilon\}}\mathbb{E}\Big|\frac{1}{|I|}\sum_{i\in I}\Phi_{Jk}(Z_{i\Delta}-Z_{(i-1)\Delta})-\int_{A(\varepsilon)}\Phi_{Jk}(x)p_{\Delta,\varepsilon}(x)\Big|,$$

where $|I|$ denotes the cardinality of the set I. To bound the last term we apply the inequality of Bretagnolle and Huber to the i.i.d. centered random variables $\big(\Phi_{Jk}(Z_{i\Delta}-Z_{(i-1)\Delta})-\mathbb{E}[\Phi_{Jk}(Z_{i\Delta}-Z_{(i-1)\Delta})]\big)_{i\in I}$ bounded by $2^{J/2+1}\|\Phi\|_\infty$, conditional to $\{\mathscr{K}_\varepsilon=I\}$. We obtain,

$$\mathbb{E}\Big|\frac{1}{|I|}\sum_{i\in I}\Phi_{Jk}(Z_{i\Delta}-Z_{(i-1)\Delta})-\int_{A(\varepsilon)}\Phi_{Jk}(x)p_{\Delta,\varepsilon}(x)\Big|\leq$$
$$C_p\sum_{k\in\Lambda_J}\Big\{\frac{1}{|I|^{p/2}}\Big[2^J\int_{A(\varepsilon)}\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\Big]^{p/2}+$$
$$\frac{2^{(p-2)(J/2+1)}\|\Phi\|_\infty^{p-2}}{|I|^{p-1}}\int_{A(\varepsilon)}2^J\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\Big\}.$$

Therefore we get

$$\sum_{k\in\Lambda_J}\mathbb{E}(|\widetilde{\alpha}_{J,k}-\alpha_{J,k}|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon)\leq C_p\sum_{k\in\Lambda_J}\Big\{\frac{1}{\widetilde{\boldsymbol{n}}(\varepsilon)^{p/2}}\Big[2^J\int_{A(\varepsilon)}\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\Big]^{p/2}$$
$$+\frac{2^{(p-2)(J/2+1)}\|\Phi\|_\infty^{p-2}}{\widetilde{\boldsymbol{n}}(\varepsilon)^{p-1}}\int_{A(\varepsilon)}2^J\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\Big\},$$

where, as developed in [26],

$$\sum_{k\in\Lambda_J}\Big[\int_{A(\varepsilon)}2^J\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\Big]^{p/2}\leq\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}2^J\|p_{\Delta,\varepsilon}\|_{L_{p/2,\varepsilon}}^{p/2}$$
$$\text{and}\quad\sum_{k\in\Lambda_J}\int_{A(\varepsilon)}2^J\Phi(2^Jx-k)^2p_{\Delta,\varepsilon}(x)dx\leq\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}2^J.$$

We can then conclude that

$$\sum_{k\in\Lambda_J}\mathbb{E}(|\widetilde{\alpha}_{J,k}-\alpha_{J,k}|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon)\leq C_p\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big\{\frac{2^J\|p_{\Delta,\varepsilon}\|_{L_{p/2,\varepsilon}}^{p/2}}{\widetilde{\boldsymbol{n}}(\varepsilon)^{p/2}}+\frac{2^{Jp/2}\|\Phi\|_\infty^{p-2}}{\widetilde{\boldsymbol{n}}(\varepsilon)^{p-1}}\Big\}.$$

Plugging this last inequality in T_5 we obtain

$$\mathbb{E}(T_5|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon)\leq C\Big\{2^{-Jsp}\Big(\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p+\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big(\frac{2^J}{\widetilde{\boldsymbol{n}}(\varepsilon)}\Big)^{p/2}\Big\},$$

where C is a constant depending on p, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, that is an upper bound of $\|p_{\Delta,\varepsilon}\|_{L_{p/2,\varepsilon}}$, and $\|\Phi\|_\infty$. Gathering all terms we get, for $p\geq 2$,

$$\mathbb{E}\Big(\|\widehat{h}_{n,\varepsilon}(\{Z_{i\Delta}-Z_{(i-1)\Delta}\}_{i\in\mathscr{K}_\varepsilon})-h_\varepsilon\|_{L_{p,\varepsilon}}^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big)$$
$$\leq C\Big\{\Big(2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p+\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big(\frac{2^J}{\widetilde{\boldsymbol{n}}(\varepsilon)}\Big)^{p/2}+(\Delta\|f\|_{L_{p,\varepsilon}})^p\Big\},$$

where C is a constant depending on p, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$ and $\|\Phi\|_\infty$. For $p\in[1,2)$, together with the additional assumption of h_ε, following the lines of the proof of Theorem 2 of Kerkyacharian and Picard [26] we obtain the same bound as above. Finally, we have established that

$$\mathbb{E}\Big(T_1|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big)\leq C\Big\{\Big(2^{-Js}\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big)^p+\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}\Big(\frac{2^J}{\widetilde{\boldsymbol{n}}(\varepsilon)}\Big)^{p/2}+(\Delta\|f\|_{L_{p,\varepsilon}})^p\Big\},\tag{31}$$

where we used that $\widetilde{\boldsymbol{n}}(\varepsilon)\leq\boldsymbol{n}(\varepsilon)$. Note that $\mathfrak{M}_\varepsilon/\lambda_\varepsilon$ remains bounded. Moreover, taking J such that $2^J=\widetilde{\boldsymbol{n}}(\varepsilon)^{\frac{1}{2s+1}}$ we have, uniformly over $\mathscr{F}(s,p,q,\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon},A(\varepsilon))$, an upper bound in $\widetilde{\boldsymbol{n}}(\varepsilon)^{-s/(2s+1)}$ for the estimation of $p_{\Delta,\varepsilon}$, which is the optimal rate of convergence for a density from $\widetilde{\boldsymbol{n}}(\varepsilon)$ direct independent observations (see [26]).

Note that we did not use that $A(\varepsilon)$ is bounded to control this quantity, it was possible to have $\overline{A}=\infty$. This together with Remark 1 lead to Proposition 5.

Control of T_3. Using the fact that Φ' is compactly supported, we get

$$\mathbb{E}(T_3|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon)\leq\frac{2^{\frac{3Jp}{2}}\|\Phi'\|_\infty^p}{\boldsymbol{n}(\varepsilon)^p}\int_{\mathbb{R}}\Big|\sum_{k\in\Lambda_J}\Phi(x-k)\Big|^p\frac{dx}{2^J}\mathbb{E}\Big(\Big|\sum_{i\in\mathscr{K}_\varepsilon}\big(M_{i\Delta}-M_{(i-1)\Delta}+b_\nu(\varepsilon)\Delta\big)\Big|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big).$$

Furthermore, we use the following upper bound for the last term in the expression above:

$$\mathbb{E}\Big(\Big|\sum_{i\in\mathscr{K}_\varepsilon}\big(M_{i\Delta}-M_{(i-1)\Delta}+b_\nu(\varepsilon)\Delta\big)\Big|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big)$$
$$\leq C_p\Big\{\mathbb{E}\Big(\Big|\sum_{i\in\mathscr{K}_\varepsilon}\big(M_{i\Delta}-M_{(i-1)\Delta}\big)\Big|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big)+\big(\widetilde{\boldsymbol{n}}(\varepsilon)b_\nu(\varepsilon)\Delta\big)^p\Big\}$$

From Rosenthal's inequality conditional on $\mathscr{I}_\varepsilon$ and $\mathscr{K}_\varepsilon$ we derive for $p\geq 2$

$$\mathbb{E}\Big(\Big|\sum_{i\in\mathscr{K}_\varepsilon}\big(M_{i\Delta}-M_{(i-1)\Delta}\big)\Big|^p|\mathscr{I}_\varepsilon,\mathscr{K}_\varepsilon\Big)\leq C_p\Big\{\widetilde{\boldsymbol{n}}(\varepsilon)\mathbb{E}(|M_\Delta|^p)+\big(\widetilde{\boldsymbol{n}}(\varepsilon)\mathbb{E}(M_\Delta^2)\big)^{\frac{p}{2}}\Big\}$$
$$=C_p\Big\{\widetilde{\boldsymbol{n}}(\varepsilon)\Delta\mu_p(\varepsilon)+\big(\widetilde{\boldsymbol{n}}(\varepsilon)\sigma^2(\varepsilon)\Delta\big)^{p/2}\Big\}$$

where $\Delta\mu_p(\varepsilon):=\mathbb{E}(|M_\Delta|^p)=O(\Delta)$. For $p\in[1,2)$ we obtain the same result using the Jensen inequality and the latter inequality with $p=2$. Next,

$$\int_{\mathbb{R}}\Big|\sum_{k\in\Lambda_J}\Phi(x-k)\Big|^p\frac{dx}{2^J}\leq 2^{-J}|\Lambda_J|^p\|\Phi\|_p^p.$$

As Φ is compactly supported, and since we estimate h_ε on an interval bounded by $\overline{A}$, for every $j \geq 0$, the set Λ_J has cardinality bounded by $|\Lambda_J| \leq C 2^J$, where C depends on the support of Φ and $\overline{A}$. It follows that,

$$\mathbb{E}(T_3 | \mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon) \leq C \|\Phi'\|_\infty^p \|\Phi\|_p^p 2^{J(5p/2-1)} \Big\{ \widetilde{\boldsymbol{n}}(\varepsilon) \boldsymbol{n}(\varepsilon)^{-p} \Delta + \boldsymbol{n}(\varepsilon)^{-p} \big(\widetilde{\boldsymbol{n}}(\varepsilon) \sigma^2(\varepsilon) \Delta \big)^{p/2} + \Big(\frac{\widetilde{\boldsymbol{n}}(\varepsilon) b_\nu(\varepsilon) \Delta}{\boldsymbol{n}(\varepsilon)} \Big)^p \Big\}. \tag{32}$$

Control of T_4. Similarly, for the last term we have

$$\mathbb{E}(T_4 | \mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon) \leq 2^{J(2p-1)} \|\Phi\|_\infty^p \|\Phi\|_p^p \Big(1 - \frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)} \Big)^p. \tag{33}$$

Deconditioning on $\mathscr{K}_\varepsilon$. Replacing (31), (32) and (33) into (30), and noticing that T_2 is negligible compared to $\mathbb{E}(T_4 | \mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon)$, we obtain

$$\begin{aligned}\mathbb{E}\big(\|\widehat{h}_{n,\varepsilon}(\{X_{i\Delta} - X_{(i-1)\Delta}\}_{i \in \mathscr{I}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p | \mathscr{I}_\varepsilon, \mathscr{K}_\varepsilon \big) &\leq C \Big\{ 2^{2Jp} \Big(1 - \frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)} \Big)^p \\ &+ \Big[\Big(2^{-Js} \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} \Big)^p + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} \Big(\frac{2^J}{\widetilde{\boldsymbol{n}}(\varepsilon)} \Big)^{p/2} + (\Delta \|f\|_{L_{p,\varepsilon}})^p \Big] \\ &+ 2^{J(3p/2-1)} |\Lambda_J|^p \Big[\widetilde{\boldsymbol{n}}(\varepsilon) \boldsymbol{n}(\varepsilon)^{-p} \Delta + \boldsymbol{n}(\varepsilon)^{-p} \big(\widetilde{\boldsymbol{n}}(\varepsilon) \sigma^2(\varepsilon) \Delta \big)^{p/2} + \Big(\frac{\widetilde{\boldsymbol{n}}(\varepsilon) b_\nu(\varepsilon) \Delta}{\boldsymbol{n}(\varepsilon)} \Big)^p \Big] \Big\}.\end{aligned}$$

where C depends on s, p, $\|h_\varepsilon\|_{L_{p,\varepsilon}}$, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. To remove the conditional expectation on $\mathscr{K}_\varepsilon$ we apply the following lemma.

Lemma 5. *Let $v_\Delta(\varepsilon) = \mathbb{P}(|M_\Delta(\varepsilon) + \Delta b_\nu(\varepsilon)| > \varepsilon)$ and $F_\Delta(\varepsilon) = \mathbb{P}(|X_\Delta| > \varepsilon)$. If $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \leq \frac{1}{3}$, then for all $r \geq 0$, there exists a constant C depending on r such that*

$$\begin{aligned}\mathbb{E}\big(\widetilde{\boldsymbol{n}}(\varepsilon)^{-r} \big| \mathscr{I}_\varepsilon \big) &\leq C \boldsymbol{n}(\varepsilon)^{-r} \\ \mathbb{E}\big((\boldsymbol{n}(\varepsilon) - \widetilde{\boldsymbol{n}}(\varepsilon))^r \big| \mathscr{I}_\varepsilon \big) &\leq C \Big\{ \Big(\boldsymbol{n}(\varepsilon) \frac{v_\Delta(\varepsilon) e^{-\lambda_\varepsilon \Delta}}{F_\Delta(\varepsilon)} \Big)^{r/2} + \Big(\boldsymbol{n}(\varepsilon) \frac{v_\Delta(\varepsilon) e^{-\lambda_\varepsilon \Delta}}{F_\Delta(\varepsilon)} \Big)^r \Big\}.\end{aligned}$$

Finally, using the fact that $\lambda_\varepsilon \Delta \to 0$ and $\widetilde{\boldsymbol{n}}(\varepsilon) \leq \boldsymbol{n}(\varepsilon)$, we conclude:

$$\begin{aligned}\mathbb{E}\big(\|\widehat{h}_{n,\varepsilon}(\{X_{i\Delta} - X_{(i-1)\Delta}\}_{i \in \mathscr{I}_\varepsilon}) - h_\varepsilon\|_{L_{p,\varepsilon}}^p | \mathscr{I}_\varepsilon \big) &\leq C \Big\{ 2^{2Jp} \Big[\Big(\frac{v_\Delta(\varepsilon)}{\boldsymbol{n}(\varepsilon) F_\Delta(\varepsilon)} \Big)^{p/2} + \Big(\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \Big)^p \Big] \\ &+ \Big[\Big(2^{-Js} \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} \Big)^p + \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon} 2^{Jp/2} \boldsymbol{n}(\varepsilon)^{-p} + (\Delta \|f\|_{L_{p,\varepsilon}})^p \Big] \\ &+ 2^{J(5p/2-1)} \Big[\boldsymbol{n}(\varepsilon)^{1-p} \Delta + \boldsymbol{n}(\varepsilon)^{-p/2} \big(\sigma^2(\varepsilon) \Delta \big)^{p/2} + (b_\nu(\varepsilon) \Delta)^p \Big] \Big\},\end{aligned}$$

where C depends on s, p, $\|h_\varepsilon\|_{L_{p,\varepsilon}}$, $\|h_\varepsilon\|_{L_{p/2,\varepsilon}}$, $\|\Phi\|_\infty$, $\|\Phi'\|_\infty$ and $\|\Phi\|_p$. The proof is now complete. $\square$

5.4 Proof of Theorem 4

Theorem 4 is a consequence of Theorem 1 and Corollary 1. For all $0 < \varepsilon \leq 1$, we decompose $\ell_{p,\varepsilon}(\widehat{f}_{n,\varepsilon}, f)$ as follows:

$$\begin{aligned}
\big[\ell_{p,\varepsilon}\big(\widehat{f}_{n,\varepsilon}, f\big)\big]^p &= \int_{A(\varepsilon)} \mathbb{E}_{P_n}\Big[\big|\widehat{\lambda}_{n,\varepsilon}\widehat{h}_{n,\varepsilon}(x) - \lambda_\varepsilon h_\varepsilon(x)\big|^p\Big]dx \\
&\leq 2^{p-1}\mathbb{E}_{P_n}\Big[\big|\widehat{\lambda}_{n,\varepsilon} - \lambda_\varepsilon\big|^p\Big]\|h_\varepsilon\|_{L_{p,\varepsilon}}^p + 2^{p-1}\mathbb{E}_{P_n}\Big[\big|\widehat{\lambda}_{n,\varepsilon}\big|^p\big\|\widehat{h}_{n,\varepsilon} - h_\varepsilon\big\|_{L_{p,\varepsilon}}^p\Big] \\
&=: 2^{p-1}(I_1 + I_2).
\end{aligned}$$

The term I_1 is controlled by means of Theorem 1 combined with the fact that if $f_\varepsilon \in \mathscr{F}(s,p,q,\mathfrak{M}_\varepsilon, A(\varepsilon))$ then $h_\varepsilon \in \mathscr{F}(s,p,q,\frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}, A(\varepsilon))$, which implies $\|h_\varepsilon\|_{L_{p,\varepsilon}} \leq \frac{\mathfrak{M}_\varepsilon}{\lambda_\varepsilon}$. Concerning the term I_2, the Cauchy-Schwarz inequality gives

$$\begin{aligned}
I_2 &= \int_{A(\varepsilon)} \mathbb{E}_{P_n}\big[\big|\widehat{\lambda}_{n,\varepsilon}\big|^p\big|\widehat{h}_{n,\varepsilon}(x) - h_\varepsilon(x)\big|^p\big]dx \\
&\leq \sqrt{\mathbb{E}_{P_n}\big[|\widehat{\lambda}_{n,\varepsilon}|^{2p}\big]}\int_{A(\varepsilon)} \sqrt{\mathbb{E}_{P_n}\big[\big|\widehat{h}_{n,\varepsilon}(x) - h_\varepsilon(x)\big|^{2p}\big]}dx = \sqrt{J_1}\sqrt{J_2}.
\end{aligned}$$

The term J_1 is treated using Theorem 1 and the triangle inequality

$$\mathbb{E}[|\widehat{\lambda}_{n,\varepsilon}|^{2p}] \leq C_p(\lambda_\varepsilon^{2p} + \mathbb{E}[|\widehat{\lambda}_{n,\varepsilon} - \lambda_\varepsilon|^{2p}]).$$

For the term J_2, notice that as $A(\varepsilon)$ is bounded, hence applying the Jensen inequality we get:

$$\int_{A(\varepsilon)} \sqrt{\mathbb{E}_{P_n}\big[\big|\widehat{h}_{n,\varepsilon}(x) - h_\varepsilon(x)\big|^{2p}\big]}dx \leq C\sqrt{\mathbb{E}_{P_n}\big[\|\widehat{h}_{n,\varepsilon} - h\|_{2p}^{2p}\big]},$$

where C depends on $\overline{A}$. The rate of the right hand side of the inequality has been studied in Corollary 1. The proof is now complete. □

6 Appendix

In this section we collect the proofs of all auxiliary results.

6.1 Proof of Lemma 1

Let $(t_n)_{n\geq 0}$ be any sequence converging to zero and let X^n be a sequence of compound Poisson processes with Lévy measure

$$\nu_n(A) := \frac{\mathbb{P}(X_{t_n} \in A)}{t_n}, \quad \forall A \in \mathscr{B}(\mathbb{R}\setminus\{0\}).$$

Using the Lévy-Khintchine formula jointly with the fact that any Lévy process has an infinitely divisible distribution we get:

$$\varphi_n(u) := \mathbb{E}\big[e^{iuX_1^n}\big] = \exp\bigg(\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu_n(dx)\bigg) = \exp\bigg(\frac{\mathbb{E}\big[e^{iuX_{t_n}}\big]-1}{t_n}\bigg)$$
$$= \exp\bigg(\frac{\big(\mathbb{E}[e^{iuX_1}]\big)^{t_n}-1}{t_n}\bigg) =: \exp\bigg(\frac{\big(\varphi(u)\big)^{t_n}-1}{t_n}\bigg). \tag{34}$$

We then deduce that the characteristic function φ_n of X_1^n converges to φ, the characteristic function of X_1, as n goes to infinity. Indeed, from (34), we have

$$\varphi_n(u) = \exp\bigg(\frac{\exp(t_n\log(\varphi(u)))-1}{t_n}\bigg) = \exp\big(\log(\varphi(u)) + O(t_n)\big).$$

This implies that the law of X_1^n converges weakly to the law of X_1.

Let us introduce the sequence of measures $\rho_n(dx) = (x^2\wedge 1)\nu_n(dx)$. From Theorem 8.7 in [34] it follows that $(\rho_n)_n$ is tight, i.e. $\sup_n \rho_n(\mathbb{R}) < \infty$ and $\lim_{\ell\to\infty}\sup_n\int_{|x|>\ell}\rho_n(dx)$. So there exists a subsequence $(\rho_{n_k})_k$ that converges weakly to a finite measure ρ. Let us introduce the measure $\widetilde{\nu}(dx) := (x^2\wedge 1)^{-1}\rho(dx)$ on $\mathbb{R}\setminus\{0\}$ and $\widetilde{\nu}(\{0\}) = 0$. Then, for any function f such that $f(x)(x^2\wedge 1)^{-1}$ is bounded, the following equalities hold:

$$\lim_{k\to\infty}\int f(x)\nu_{n_k}(dx) = \lim_{k\to\infty}\int f(x)(x^2\wedge 1)^{-1}\rho_{n_k}(dx)$$
$$= \int f(x)(x^2\wedge 1)^{-1}\rho(dx) = \int f(x)\widetilde{\nu}(dx)$$

By definition of ν_n, this implies that

$$\lim_{n\to\infty}\frac{1}{t_n}\mathbb{E}\big[f(X_{t_n})\big] = \int f(x)\widetilde{\nu}(dx).$$

The uniqueness of $\widetilde{\nu}$ (see [34], p. 43) joint with the fact that ν_n converges weakly to ν (since the law of X_1^n converges weakly to the law of X_1), allow us to conclude that $\widetilde{\nu}\equiv\nu$. $\square$

6.2 Proof of Lemma 3

Using the definition (22) we derive that

$$p_{\Delta,\varepsilon} - h_\varepsilon = h_\varepsilon\Big(\frac{e^{-\lambda_\varepsilon\Delta}\lambda_\varepsilon\Delta}{1-e^{-\lambda_\varepsilon\Delta}} - 1\Big) + \sum_{k=2}^{\infty}\frac{e^{-\lambda_\varepsilon\Delta}(\lambda_\varepsilon\Delta)^k}{k!(1-e^{-\lambda_\varepsilon\Delta})}h_\varepsilon^{\star k}$$
$$= \Big(\frac{e^{-\lambda_\varepsilon\Delta}\lambda_\varepsilon\Delta}{1-e^{-\lambda_\varepsilon\Delta}} - 1\Big)h_\varepsilon + (\lambda_\varepsilon\Delta)^2\sum_{k=2}^{\infty}\frac{e^{-\lambda_\varepsilon\Delta}(\lambda_\varepsilon\Delta)^{k-2}}{k!(1-e^{-\lambda_\varepsilon\Delta})}h_\varepsilon^{\star k}.$$

Taking the L_p norm and using the Young inequality together with the fact that h_ε is a density with respect to the Lebesgue measure, i.e. $\|h_\varepsilon\|_{L_{1,\varepsilon}} \le 1$, we get

$$\big\|p_{\Delta,\varepsilon} - h_\varepsilon\big\|_{L_{p,\varepsilon}} \le \Big(\Big|\frac{e^{-\lambda_\varepsilon\Delta}\lambda_\varepsilon\Delta}{1-e^{-\lambda_\varepsilon\Delta}} - 1\Big| + \frac{(\lambda_\varepsilon\Delta)^2}{1-e^{-\lambda_\varepsilon\Delta}}\Big)\|h_\varepsilon\|_{L_{p,\varepsilon}}.$$

Under the regime $\lambda_\varepsilon\Delta \to 0$ we obtain

$$\big\|p_{\Delta,\varepsilon} - h_\varepsilon\big\|_{L_{p,\varepsilon}} \le C\lambda_\varepsilon\Delta\|h_\varepsilon\|_{L_{p,\varepsilon}},$$

where $C > 2$ and $\lambda_\varepsilon h_\varepsilon = f_\varepsilon$, which completes the proof. □

6.3 Proof of Lemma 4

We have

$$\boldsymbol{n}(\varepsilon) = \sum_{i=1}^{n} \mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta} - X_{(i-1)\Delta}|) = \widehat{\lambda}_{n,\varepsilon} n\Delta.$$

We introduce the centered i.i.d. random variables $V_i = \mathbb{1}_{(\varepsilon,\infty)}(|X_{i\Delta} - X_{(i-1)\Delta}|) - F_\Delta(\varepsilon)$, which are bounded by 2 and such that $\mathbb{E}[V_i^2] \le F_\Delta(\varepsilon)$. Applying the Bernstein inequality we have,

$$\mathbb{P}\Big(\Big|\frac{\boldsymbol{n}(\varepsilon)}{n} - F_\Delta(\varepsilon)\Big| > x\Big) \le 2\exp\Big(-\frac{nx^2}{2(F_\Delta(\varepsilon) + \frac{2x}{3})}\Big), \quad x > 0. \tag{35}$$

Fix $x = F_\Delta(\varepsilon)/2$, on the set $A_x = \big\{\big|\frac{\boldsymbol{n}(\varepsilon)}{n} - F_\Delta(\varepsilon)\big| \le x\big\}$ we have

$$n\frac{F_\Delta(\varepsilon)}{2} \le \boldsymbol{n}(\varepsilon) \le n\frac{3F_\Delta(\varepsilon)}{2}. \tag{36}$$

Moreover it holds that

$$\mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\big) = \mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\mathbb{1}_{A_x^c}\big) + \mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\mathbb{1}_{A_x}\big).$$

Since $r \ge 0$ and $\boldsymbol{n}(\varepsilon) \ge 1$, using (35) and (36) we get the following upper bound

$$\mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\big) \le 2\exp\big(-\tfrac{3}{32}nF_\Delta(\varepsilon)\big) + \Big(\frac{nF_\Delta(\varepsilon)}{2}\Big)^{-r}$$

and the lower bound

$$\mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\big) \ge \mathbb{E}\big(\boldsymbol{n}(\varepsilon)^{-r}\mathbb{1}_{A_x}\big) \ge \Big(\frac{3nF_\Delta(\varepsilon)}{2}\Big)^{-r}.$$

This completes the proof. □

6.4 Proof of Lemma 5

For the first inequality, the proof is similar to the proof of Lemma 4. Using the definition of $\widetilde{\boldsymbol{n}}(\varepsilon)$ we have

$$\widetilde{\boldsymbol{n}}(\varepsilon) = \sum_{i\in\mathscr{I}_\varepsilon} \mathbb{I}_{Z_{i\Delta}(\varepsilon)\neq Z_{(i-1)\Delta}(\varepsilon)}.$$

For $i \in \mathscr{I}_\varepsilon$, we set $W_i := \mathbb{I}_{Z_{i\Delta}(\varepsilon)\neq Z_{(i-1)\Delta}(\varepsilon)}$. We have

$$\mathbb{E}(W_i|i \in \mathscr{I}_\varepsilon) = \mathbb{P}\big(Z_{i\Delta}(\varepsilon) \neq Z_{(i-1)\Delta}(\varepsilon)\big||X_{i\Delta} - X_{(i-1)\Delta}| > \varepsilon\big) = 1 - \frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}$$

using the independence of $M(\varepsilon)$ and $Z(\varepsilon)$. The variables $W_i - \mathbb{E}(W_i|i \in \mathscr{I}_\varepsilon)$ are centered i.i.d., bounded by 2 and such that the following bound on the variance holds: $\mathbb{V}(W_i|\mathscr{I}_\varepsilon) \leq \frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}$. Applying the Bernstein inequality we have,

$$\mathbb{P}\Big(\Big|\frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)} - \Big(1 - \frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)\Big| > x\Big|\mathscr{I}_\varepsilon\Big) \leq 2\exp\Big(-\frac{\boldsymbol{n}(\varepsilon)x^2}{2\big(\frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)} + \frac{2x}{3}\big)}\Big), \quad x > 0. \quad (37)$$

Fix $x = \frac{1}{2}$, on the set $A_x = \big\{\big|\frac{\widetilde{\boldsymbol{n}}(\varepsilon)}{\boldsymbol{n}(\varepsilon)} - (1 - \frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)})\big| \leq \frac{1}{2}\big\}$ we have

$$\frac{\boldsymbol{n}(\varepsilon)}{6} < \Big(\frac{1}{2} - \frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)\boldsymbol{n}(\varepsilon) \leq \widetilde{\boldsymbol{n}}(\varepsilon), \quad (38)$$

if $\frac{v_\Delta(\varepsilon)}{F_\Delta(\varepsilon)} \leq \frac{1}{3}$. It follows from (37), (38) and $\widetilde{\boldsymbol{n}}(\varepsilon) \geq 1$ that for $r \geq 0$

$$\mathbb{E}\big(\widetilde{\boldsymbol{n}}(\varepsilon)^{-r}\big|\mathscr{I}_\varepsilon\big) \leq 2\exp\big(-\tfrac{3}{16}\boldsymbol{n}(\varepsilon)\big) + \Big(\frac{\boldsymbol{n}(\varepsilon)}{6}\Big)^{-r}.$$

Finally, using that for all $x > 0$ we have $x^r e^{-x} \leq C_r := r^r e^{-r}$ we derive

$$\mathbb{E}\big(\widetilde{\boldsymbol{n}}(\varepsilon)^{-r}\big|\mathscr{I}_\varepsilon\big) \leq C_r\boldsymbol{n}(\varepsilon)^{-r} + \Big(\frac{\boldsymbol{n}(\varepsilon)}{6}\Big)^{-r},$$

which leads to the first part of the result.

The second part of the result can be obtained by means of Rosenthal's inequality. For $r \geq 0$, we have, using that $\boldsymbol{n}(\varepsilon) \geq \widetilde{\boldsymbol{n}}(\varepsilon)$,

$$\mathbb{E}\big((\boldsymbol{n}(\varepsilon) - \widetilde{\boldsymbol{n}}(\varepsilon))^r\big|\mathscr{I}_\varepsilon\big) \leq C_r\Big\{\mathbb{E}\Big(\Big|\boldsymbol{n}(\varepsilon)\big(1 - \tfrac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\big) - \widetilde{\boldsymbol{n}}(\varepsilon)\Big|^r\Big|\mathscr{I}_\varepsilon\Big) + \Big(\boldsymbol{n}(\varepsilon)\tfrac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)^r\Big\}.$$

The Rosenthal inequality leads to, for $r \geq 2$,

$$E\Big(\Big|\boldsymbol{n}(\varepsilon)\big(1 - \tfrac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\big) - \widetilde{\boldsymbol{n}}(\varepsilon)\Big|^r\Big|\mathscr{I}_\varepsilon\Big) \leq C_r\Big(\boldsymbol{n}(\varepsilon)\tfrac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)^{r/2}.$$

Thanks to Jensen's inequality we can also treat the case $0 < r < 2$ recovering the same inequality. Therefore, it follows that for all $r > 0$

$$\mathbb{E}\big((\boldsymbol{n}(\varepsilon) - \widetilde{\boldsymbol{n}}(\varepsilon))^r\big|\mathscr{I}_\varepsilon\big) \leq C\Big\{\Big(\boldsymbol{n}(\varepsilon)\frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)^{r/2} + \Big(\boldsymbol{n}(\varepsilon)\frac{v_\Delta(\varepsilon)e^{-\lambda_\varepsilon\Delta}}{F_\Delta(\varepsilon)}\Big)^r\Big\}.$$

This completes the proof. □

References

[1] O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick. *Lévy processes: theory and applications*. Springer Science & Business Media, 2012.

[2] M. Bec and C. Lacour. “Adaptive pointwise estimation for pure jump Lévy processes”. In: *Statistical Inference for Stochastic Processes* 18.3 (2015), pp. 229–256.

[3] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, and M. Reiß. *Lévy Matters IV - Estimation for discretely observed Lévy processes*. Vol. 2128. Springer International Publishing, 2015.

[4] F. Biagini, Y. Bregman, and T. Meyer-Brandis. *Electricity futures price modeling with Lévy term structure models*. Tech. rep. LMU München Working Paper, 2011.

[5] O. Boxma, J. Ivanovs, K. Kosiński, and M. Mandjes. “Lévy-driven polling systems and continuous-state branching processes”. In: *Stochastic Systems* 1.2 (2011), pp. 411–436.

[6] B. Buchmann and R. Grübel. “Decompounding: an estimation problem for Poisson random sums”. In: *Ann. Statist.* 31.4 (2003), pp. 1054–1074.

[7] P. Carr, H. Geman, D. B. Madan, and M. Yor. “The fine structure of asset returns: An empirical investigation*”. In: *The Journal of Business* 75.2 (2002), pp. 305–333.

[8] A. Coca. “Efficient nonparametric inference for discretely observed compound Poisson processes”. In: *Probab. Theory Related Fields* (To appear).

[9] F. Comte and V. Genon-Catalot. “Nonparametric estimation for pure jump Lévy processes based on high frequency data”. In: *Stochastic Process. Appl.* 119.12 (2009), pp. 4088–4123.

[10] F. Comte, C. Duval, and V. Genon-Catalot. “Nonparametric density estimation in compound Poisson processes using convolution power estimators”. In: *Metrika* 77.1 (2014), pp. 163–183.

[11] F. Comte and V. Genon-Catalot. “Estimation for Lévy processes from high frequency data within a long time interval”. In: *Ann. Statist.* 39.2 (2011), pp. 803–837.

[12] F. Comte and V. Genon-Catalot. “Nonparametric adaptive estimation for pure jump Lévy processes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 46.3 (2010), pp. 595–617.

[13] R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004, xvi+535.

[14] C. Duval. “Density estimation for compound Poisson processes from discrete data”. In: *Stochastic Process. Appl.* 123.11 (2013), pp. 3963–3986.

[15] B. van Es, S. Gugushvili, and P. Spreij. “A kernel type nonparametric density estimator for decompounding”. In: *Bernoulli* 13.3 (2007), pp. 672–694.

[16] J. E. Figueroa-López. “Nonparametric estimation of Lévy models based on discrete-sampling”. In: *Optimality*. Vol. 57. IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH, 2009, pp. 117–146.

[17] J. E. Figueroa-López. “Sieve-based confidence intervals and bands for Lévy densities”. In: *Bernoulli* 17.2 (2011), pp. 643–670.

[18] J. E. Figueroa-López and C. Houdré. “Small-time expansions for the transition distributions of Lévy processes”. In: *Stochastic Process. Appl.* 119.11 (2009), pp. 3862–3889.

[19] J Gil-Pelaez. “Note on the inversion theorem”. In: *Biometrika* 38.3-4 (1951), pp. 481–482.

[20] S. Gugushvili. “Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process”. In: *Journal of Nonparametric Statistics* 21.3 (2009), pp. 321–343.

[21] S. Gugushvili. “Nonparametric inference for discretely sampled Lévy processes”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 48.1 (2012), pp. 282–307.

[22] S. Gugushvili, F. van der Meulen, and P. Spreij. “Nonparametric Bayesian inference for multidimensional compound Poisson processes”. In: *arXiv preprint arXiv:1412.7739* (2014).

[23] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*. Vol. 129. Springer Science & Business Media, 2012.

[24] J. Kappus. “Adaptive nonparametric estimation for Lévy processes observed at low frequency”. In: *Stochastic Process. Appl.* 124.1 (2014), pp. 730–758.

[25] J. Kappus and M. Reiß. “Estimation of the characteristics of a Lévy process observed at arbitrary frequency”. In: *Stat. Neerl.* 64.3 (2010), pp. 314–328.

[26] G. Kerkyacharian and D. Picard. “Density estimation in Besov spaces”. In: *Statistics & Probability Letters* 13.1 (1992), pp. 15–24.

[27] V. Konakov and V. Panov. “Sup-norm convergence rates for Lévy density estimation”. In: *Extremes* 19.3 (2016), pp. 371–403.

[28] E. Mariucci. “Asymptotic equivalence for pure jump Lévy processes with unknown Lévy density and Gaussian white noise”. In: *Stochastic Process. Appl.* 126.2 (2016), pp. 503–541.

[29] M. H. Neumann and M. Reiß. “Nonparametric estimation for Lévy processes from low-frequency observations”. In: *Bernoulli* 15.1 (2009), pp. 223–248.

[30] R. Nickl and M. Reiß. “A Donsker theorem for Lévy measures”. In: *J. Funct. Anal.* 263.10 (2012), pp. 3306–3332.

[31] R. Nickl, M. Reiß, J. Söhl, and M. Trabs. “High-frequency Donsker theorems for Lévy measures”. In: *Probab. Theory Related Fields* 164.1-2 (2016), pp. 61–108.

[32] R. C. Noven, A. E. Veraart, and A. Gandy. “A Lévy-driven rainfall model with applications to futures pricing”. In: *AStA Advances in Statistical Analysis* 99.4 (2015), pp. 403–432.

[33] L. Rüschendorf and J. H. C. Woerner. “Expansion of transition distributions of Lévy processes in small time”. In: *Bernoulli* 8.1 (2002), pp. 81–96.

[34] K.-i. Sato. *Lévy processes and infinitely divisible distributions*. Vol. 68. Cambridge Studies in Advanced Mathematics. Translated from the 1990 Japanese original, Revised by the author. Cambridge: Cambridge University Press, 1999, pp. xii+486.

[35] M. Trabs. “Quantile estimation for Lévy measures”. In: *Stochastic Process. Appl.* 125.9 (2015), pp. 3484–3521.