Nonparametric estimation for irregularly sampled Lévy processes

Johanna Kappus

Institut für Mathematik
Universität Rostock

May 27, 2020

Abstract

We consider nonparametric statistical inference for Lévy processes sampled irregularly, at low frequency. The estimation of the jump dynamics as well as the estimation of the distributional density are investigated.
Non-asymptotic risk bounds are derived and the corresponding rates of convergence are discussed under global as well as local regularity assumptions. Moreover, minimax optimality is proved for the estimator of the jump measure. Some numerical examples are given to illustrate the practical performance of the estimation procedure.

Keywords. Nonparametric statistical inference. Lévy processes. Irregular sampling. Density estimation.
AMS Subject Classification. 62G05 62G07 62G20 62M05

1 Introduction

Nonparametric statistical inference for stochastic processes with jumps has a long history, dating back as far as to the work by Rubin and Tucker (1959) or Basawa and Brockwell (1982).
In the past decade, jump processes have become increasingly popular among practitioners, especially in the field of financial applications, and the interest in the topic has constantly grown. In particular, a vast amount of literature has been published on the estimation of the characteristics of Lévy processes.
In estimation problems for Lévy processes, one can essentially distinguish between two different types of observation schemes: In a high frequency framework, it is assumed that the maximal distance between the observation times tends to zero as the number n of observations increases to infinity, whereas in a low frequency regime, it is not assumed that the observation distances are asymptotically small.
So far, when low frequent observations of the underlying Lévy process are considered, most publications have focused on the case where the observations are homogeneous, which means that the distance Δ between any two observation times t_j and t_{j+1} is fixed an does not vary with j, see for example, Neumann and Reiß (2009), Comte and Genon-Catalot (2010) or Kappus (2014).
Contrarily to this, when a high frequent sampling scheme is being investigated, the homogeneity assumption can easily be disposed of, see, among many others, the work by Figueroa-López (2009) or Comte and Genon-Catalot (2009, 2011).
In the present work, we focus on nonparametric statistical inference for Lévy processes when the sampling scheme is low frequent and irregular. This means that the (deterministic) distances Δ_j between any two observation times t_j and t_{j+1} may vary in j, from very small values close to zero to some possibly very large value Δ_{max}.
Belomestny (2011) and, most recently, Belomestny and Trabs (2015) have investigated the estimation of the characteristics of a Lévy process (X_t), when homogeneous and low frequent observation of a time changed process $Y_t = X_{\mathcal{T}(t)}$ are available. In this framework $\mathcal{T}$ is understood to be a random time change, independent of (X_t).

It is important to point out that the deterministic and irregular sampling scheme discussed in the present work is not embedded, as as special case, in the time changed model which has been investigated in Belomestny (2011). In that paper, the estimator is constructed for a stationary time-change process and under the standing assumption, referred to as (ATI), that $\mathbb{E}[\mathcal{T}(t)] = t$. However, when a deterministic time change is considered, this would readily imply that $\mathcal{T}(t) = t$ and the problem would hence degenerate to a standard estimation problem with homogeneous observations.
The statistical framework considered in the present publication, with arbitrary irregular sampling, is hence new in the literature on nonparametric estimation for Lévy processes. We focus on the following problems: Firstly, the nonparametric estimation of the jump measure is being discussed under some additional assumptions on the process. Secondly, we discuss the estimation of the distributional density of X_1 under very general a priori assumptions on the process. This second problem has not yet received much attention in the literature on Lévy processes and is, indeed, quite standard when high frequent or heterogeneous observations are available. However, under low frequent and irregular sampling, the estimation of distributional densities is not straightforward.
This paper is organized as follow: In Section , the statistical framework is made precise and some notation is introduced. In Section , the estimation of the jump measure is investigated. An oracle type estimator for the Lévy density is introduced, which turns out to depend on the appropriate choice of certain weight functions. Non-asymptotic bounds are derived for the oracle estimator and corresponding rates of convergence are derived and shown to be optimal in a minimax sense. The density estimation is then addressed in Section . In Section , an algorithm for the fully data driven choice of the weights as well as for the choice of the cutoff parameter is introduced. All proofs are postponed to Section .

2 Statistical model, assumptions and notation

A one dimensional Lévy process $X = \{X_t : t \geq 0\}$ is observed at deterministic time points $0 = t_0 < \cdots < t_n =: T$. Throughout the rest of this work, we assume that there exists a positive constant $\Delta_{\max}$ such that $\forall j \in \mathbb{N},\ t_j - t_{j-1} \leq \Delta_{\max}$. Apart from this upper bound, no additional assumptions on the observation times are imposed so the sampling is irregular and fully general. The goal of this paper is twofold. Firstly, we focus on the estimation of the jump dynamics under the following additional assumptions.

(A1) X has finite variation on compact sets.

(A2) X has no drift component.

(A3) For one and hence for any $t > 0$, $\mathbb{E}[X_t^2] < \infty$.

(A4) The Lévy measure ν has a Lebesgue density η which is continuous on $\mathbb{R}\setminus\{0\}$.

Under (A1)-(A4), the characteristic function of X_Δ is known to admit the following representation.

$$\varphi_\Delta(u) := \mathbb{E}\left[e^{iuX_\Delta}\right] = e^{\Delta\Psi(u)}, \tag{2.1}$$

with characteristic exponent

$$\Psi(u) = \int \left(e^{iux} - 1\right)\eta(x)\,\mathrm{d}x = \int \frac{e^{iux} - 1}{x} x\eta(x)\,\mathrm{d}x. \tag{2.2}$$

For the proof, see e.g. Theorem 8.1 in Sato (1999). The process is then entirely described by the jump measure and hence by the function $g(x) := x\eta(x)$. We discuss the estimation of g, with $\mathrm{L}^2(\Omega)-$loss on some interval $\Omega := [\omega_1, \omega_2]$, $-\infty \leq \omega_1 < \omega_2 \leq \infty$.

The above assumptions are met for many prototypical Lévy processes such as compound Poisson processes, gamma processes or tempered stable processes without drift and with index of stability $\alpha \in (0,1)$. Notice that $g \in \mathrm{L}^2(\mathbb{R})$ may fail to hold true, but is satisfied if $\eta(x) = O(|x|^{-\alpha}),\ |x| \to 0$ for some $\alpha < \frac{3}{2}$. This is met, for example, for compound Poisson processes, gamma-processes and tempered stable processes with $\alpha \in (0, 1/2)$. If $g \notin \mathrm{L}^2(\mathbb{R})$, estimating g with $\mathrm{L}^2(\Omega)$-loss does still make sense for compact sets Ω bounded away from the origin.
It is worth mentioning that the assumptions (A1) and (A2) can be omitted and a fully general treatment of the problem is possible, but at the cost of additional technical complications, see, for example, Neumann and Reiß (2009) for the homogeneous case.
Secondly, we investigate the estimation of the distributional density f of X_1 with $\mathrm{L}^2(\mathbb{R})$-loss, in case that a square integrable Lebesgue density exists. It is well known that X_1 possesses a density $f \in \mathrm{L}^2(\mathbb{R})$ if the process has a high enough activity of small jumps or a non-zero Gaussian component. This follows, for example, from arguments given in Orey (1968). For example, the Lebesgue density exists and is square integrable if the process is of pure jump type and there exist constants $C > 0$ and $\beta > \frac{1}{2}$ such that $\eta(x) + \eta(-x) > \beta|x|^{-1}\ \forall x \in [0, C]$.
Typical examples are gamma processes with scale parameter $\beta > \frac{1}{2}$, stable or tempered stable processes as well as the Brownian motion. Prototypic counterexamples are compound Poisson processes or gamma processes with $\beta \leq \frac{1}{2}$.
We conclude this section by introducing some notation which will be used throughout the rest of the text.
For a function $f \in \mathrm{L}^1(\mathbb{R}) \cup \mathrm{L}^2(\mathbb{R})$, $\mathcal{F}f$ is understood to be the Fourier transform. By Δ_j, we denote the distance $t_j - t_{j-1}$ between the observation times. $Z_j = X_j - X_{j-1}$ is the corresponding increment of the process and

$$\varphi_j(u) := \mathbb{E}\big[e^{iuZ_j}\big] = \mathbb{E}\big[e^{iuX_{\Delta_j}}\big]$$

its characteristic function. Given a kernel K and bandwidth $h > 0$, we write $\mathrm{K}_h(x) := 1/h\,\mathrm{K}(x/h)$. Moreover, $\Delta_{\mathrm{Max}} := \max\{\Delta_{max}, 1\}$. Let X_+ and X_- be independent Lévy processes with Lévy measures $\nu_+(\mathrm{d}x) = \nu|_{(0,\infty)}(\mathrm{d}x)$ and $\nu_-(\mathrm{d}x) := \nu|_{(-\infty,0)}(\mathrm{d}(\text{-}x))$. For $m \in \mathbb{N}$, $C_m^{\pm} := \mathbb{E}[X_+^m] + \mathbb{E}[X_-^m]$. Finally, given a continuous function f and $u \in \mathbb{R}$

$$\|f\|_{\mathrm{L}^k,u}^k := \int_{-u}^{u} |f(z)|^k \,\mathrm{d}z \quad \text{and} \quad \|f\|_{\infty,u} := \sup_{z\in[-u,u]} |f(z)|.$$

3 Estimation of the jump dynamics

3.1 Estimation procedure and non-asymptotic risk bounds

It follows from formula () and formula () that the Fourier transform of g can be recovered by differentiating the characteristic exponent,

$$\Psi'(u) = \int \frac{\mathrm{d}}{\mathrm{d}u} e^{iux} \nu(\mathrm{d}x) = i \int e^{iux} x\eta(x)\,\mathrm{d}x\,\mathrm{d}x = i\mathcal{F}g(u). \tag{3.1}$$

For (possibly complex) weight functions $w_1, \cdots, w_n$ to be chosen appropriately , we define

$$p(u) := \sum_{j=1}^{n} w_j(u)\varphi'_{\Delta_j}(u) = \sum_{j=1}^{n} w_j(u)\Delta_j\Psi'(u)\varphi_{\Delta_j}(u) \quad \text{and} \quad q(u) := \sum_{j=1}^{n} w_j(u)\Delta_j\varphi_{\Delta_j}(u).$$

The derivative of the characteristic exponent then equals the ratio p/q and formula () permits to recover the Fourier transform of g as follows,

$$\mathcal{F}g(u) = \Psi'(u)/i = \frac{p(u)}{iq(u)}.$$

Let us introduce the empirically accessible counterparts of p and q,

$$\widehat{p}(u) := \sum_{j=1}^{n} w_j(u)\widehat{\varphi}_j'(u) := \sum_{j=1}^{n} w_j(u) i Z_j e^{iuZ_j}$$

and

$$\widehat{q}(u) := \sum_{j=1}^{n} \Delta_j w_j(u)\widehat{\varphi}_j(u) := \sum_{j=1}^{n} \Delta_j w_j(u) e^{iuZ_j}.$$

Taking expectation, we find that these quantities are unbiased estimators of p and q. This suggests to use $\widehat{p}/\widehat{q}$ as an estimator of Ψ'.
However, the estimation of the Lévy density from low frequent observations is a prototypical statistical inverse problem and the rates of convergence are governed by the smoothness of the underlying density as well as the decay behavior of the function q in the denominator. Too small values of the denominator will typically lead to a highly irregular behavior of the estimator and hence a large variance. Inspired by Neumann (1997), we introduce a regularized version of the inverse of $\widehat{q}$. Once $\widehat{q}$ is below some threshold which is of the order of the standard deviation, a reasonable estimate of $1/q$ is no longer possible so the estimator is set to zero. Moreover, in order to ensure that the estimator is bounded from below, we introduce an additional constant threshold value. For some threshold parameter $\kappa > 0$ to be chosen,

$$\frac{1}{\widetilde{q}(u)} := \frac{1\left(\{|\widehat{q}(u)| \geq \max\{\sigma(u), \kappa\}\}\right)}{\widehat{q}(u)}, \quad \text{with} \quad \sigma(u) := \sqrt{\sum_{j=1}^{n} \Delta_j^2 |w_j(u)|^2}\,.$$

The corresponding regularized estimator of Ψ' is $\widehat{\Psi'}(u) := \widehat{p}/\widetilde{q}$. Finally, let K be a kernel function having an integrable Fourier transform. For $h > 0$, the corresponding kernel estimator of g is

$$\widehat{g}_h(x) := \mathcal{F}^{-1}\left(\mathcal{F}\mathrm{K}_h \frac{\widehat{p}}{i\widetilde{q}}\right)(x) = \frac{1}{2\pi}\int e^{-iux}\mathcal{F}\mathrm{K}(hu)\frac{\widehat{p}(u)}{i\widetilde{q}(u)}\,\mathrm{d}u.$$

There remains to specify the w_j. It is intuitive that the optimal choice of w_j will depend on Δ_j as well as on φ_j. For small values of φ_j, the noise dominates so the quality of the estimator gets worse, which should lead to choosing a relatively small weight. Moreover, the smaller Δ_j is, the more information on the jump dynamics is contained in the observation, which motivates to give a high weight to $\widehat{\varphi}_j$. These considerations lead to specifying the ideal (oracle) weights $w_j^*(u) := \overline{\varphi_{\Delta_j}(u)}$. Notice that this statistical framework has strong structural similarities to a deconvolution problem with heteroscedastic errors, see Delaigle and Meister (2008).
In what follows, $\widehat{g}_h$ is understood to be the oracle estimator corresponding to the ideal weights $w_j = w_j^*$, $j = 1, \cdots, n$. It is clear, however, that the w_j^* are not feasible to actually compute, since they depend on the (unknown) characteristic function. In the sequel, we derive risk bounds and optimality properties for the oracle estimator $\widehat{g}_h$. A procedure for the fully data driven choice of the weights is then proposed in Section .

Theorem 3.1. *Assume that $\mathbb{E}\left[|X_1|^4\right] < \infty$. Assume, moreover, that $g|_\Omega \in \mathrm{L}^2(\Omega)$. Then there exists some $C : \mathbb{R}_+^2 \to \mathbb{R}_+$ which is monotonously increasing with respect to both components such that*

$$\begin{aligned}\mathbb{E}\left[\|\widehat{g}_{h,n} - g\|^2_{\mathrm{L}^2(\Omega)}\right] \leq & 2\|g - \mathrm{K}_h * g\|^2_{\mathrm{L}^2(\Omega)} + C(\Delta_{Max}, C_4^{\pm}) \int \frac{|\mathcal{F}\,\mathrm{K}(hu)|^2}{q(u)}\, du \\ = & 2\|g - \mathrm{K}_h * g\|^2_{\mathrm{L}^2(\Omega)} + C(\Delta_{Max}, C_4^{\pm}) \int \frac{|\mathcal{F}\,\mathrm{K}(hu)|^2}{\sum_{j=1}^n \Delta_j |\varphi_{\Delta_j}(u)|^2}\, du.\end{aligned}$$

Discussion. It is interesting to note that the upper risk bound in the preceding theorem confirms the analogy of the nonparametric estimation of g with a deconvolution problem with heteroscedastic errors, see Delaigle and Meister (2008).
We concentrate, in this work, on a deterministic sampling scheme. However, the construction of the estimator and the upper bounds may be generalized to the case where the process is sampled at random times, provided that the random sampling does not depend on the underlying process X. In the proof of the rate results, one will then have to consider conditional expectations with respect to $\mathcal{T}$.

3.2 Rates of convergence

3.2.1 Minimax upper bounds

In this section, we study the asymptotic properties of $\widehat{g}_h$ which can be derived from the upper risk bound formulated in Theorem 3.1.
Global regularity

We start by considering the estimation of g with L^2-loss on the whole real axis and the resulting rates of convergence over certain nonparametric classes. It has been pointed out that the estimation of g resembles the estimation of a distributional density from observations with additional heteroscedastic errors. The rates of convergence will thus depend on the Δ_j and on the decay of the $|\varphi_j| = |\varphi|^{\Delta_j}$, as well as on the smoothness of g.
However, one needs to beware of the fact that (unlike in a standard deconvolution framework) the smoothness of g and the decay of φ are not independent of each other.
It is easily seen that the polynomial or exponential decay conditions

$$|\varphi(u)| \sim |u|^{-\beta},\ u \to \infty \quad \text{or} \quad |\varphi(u)| \sim \exp(-c|u|^{\alpha})$$

imply a logarithmic or polynomial growth of the characteristic exponent which gives, in turn,

$$\limsup_{u\to\infty} \frac{|\mathcal{F}g(u)|}{|u|^{-1}} = \limsup_{u\to\infty} \frac{|\Psi'(u)|}{|u|^{-1}} > 0 \quad \text{or} \quad \limsup_{u\to\infty} \frac{|\mathcal{F}g(u)|}{|u|^{\alpha-1}} > 0.$$

The faster the characteristic function φ of X_1 decays, the slower will hence be the decay of the Fourier transform of g. For processes of compound Poisson type, the absolute value of φ is bounded from below and g may be infinitely differentiable.

Let us introduce the following nonparametric classes of functions and of corresponding Lévy processes: For $\beta > 0$ and $\alpha \in (0, 1/2)$, let $\mathcal{G}_{\mathrm{pol}}(\beta, C_\varphi, c_\varphi, C_g, C)$ or $\mathcal{G}_{\exp}(\alpha, C_\varphi, C_g, C)$ be the classes of functions $g(x) = x\eta(x)$ such that for the corresponding Lévy process, (A1)-(A4) are met, $C_4^{\pm} < C$,

$$\forall u \in \mathbb{R}:\ |\varphi_{X_1}(u)| \geq (1 + C_\varphi |u|^2)^{-\frac{\beta}{2}} \quad \text{or} \quad \forall u \in \mathbb{R}:\ |\varphi_{X_1}(u)| \geq C_\varphi \exp(-c_\varphi |u|^{\alpha}),$$

and, in addition,

$$\forall u \in \mathbb{R}: |\mathcal{F}g(u)| \leq C_g |u|^{-1} \quad \text{or} \quad \forall u \in \mathbb{R}:\ |\mathcal{F}g(u)| \leq C_g |u|^{\alpha-1}.$$

Moreover, let $\mathcal{G}_{\mathrm{cp}}(C_\varphi, a, \rho, C_g, c_g, C)$ be the class of functions g such that X is a compound Poisson process, $C_4^{\pm} < C$,

$$\forall u \in \mathbb{R} : \ |\varphi_{X_1}(u)| \geq C_\varphi$$

and

$$\forall u \in \mathbb{R} : |\mathcal{F}g(u)| \leq C_g |u|^{-a} \exp(-c_g |u|^\rho).$$

The following result which describes the rates of convergence with respect to the prescribed smoothness classes introduced above is a direct consequence of Theorem 3.1.

Proposition 3.2. *Let* K *be the sinc-kernel. This is equivalent to stating that* $\mathcal{F}\mathrm{K}(u) = \mathbb{1}_{[-1,1]}(u)$.

(i) Let h^* be implicitly defined, as the solution of the minimization equation

$$\sum_{j=1}^n \Delta_j h^{2\Delta_j \beta + 2} = 1. \tag{3.2}$$

Then

$$\sup_{g \in \mathcal{G}_{pol}(\beta, C_\varphi, c_\varphi, C_g, C)} \mathbb{E}_g \big[\|g - \widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})} \big] = O(h^*).$$

(ii) Let h^* be the solution of

$$\sum_{j=1}^n \Delta_j e^{-2\Delta_j c_\varphi (1/h)^\alpha} h^{2(\alpha-1)} = 1.$$

Then

$$\sup_{g \in \mathcal{G}_{exp}(\alpha, C_\varphi, C_g, C)} \mathbb{E}_g \big[\|g - \widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})} \big] = O\big((h^*)^{1-2\alpha}\big).$$

(iii) Let h^* be the solution of

$$e^{2c_g (1/h)^\rho} h^{-2a} = \sum_{j=1}^n \Delta_j C_\varphi^{\Delta_j}.$$

a) Assume that $\rho > 0$ holds and $a = (1-\rho)/2$. Then

$$\sup_{g \in \mathcal{G}_{cp}(C_\varphi, a, \rho, C_g, c_g, C)} \mathbb{E}_g \big[\|g - \widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})} \big] = O\big(\exp(-2c_g (1/h^*)^\rho)\big).$$

b) Assume that $\rho = 0$ and $a > 0$. Then

$$\sup_{g\in\mathcal{G}_{cp}(C_\varphi,a,\rho,C_g,c_g,C)} \mathbb{E}_g\big[\|g-\widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})}\big] = O\big((h^*)^{2a-1}\big).$$

The convergence rates summarized above, with h^* defined implicitly, are not particularly intuitive. For this reason, we give some examples to better understand the underlying structure.

Examples.

(i) Consider the special case of homogeneous observations, $\Delta_j = \Delta,\ j = 1,...,n$. Then, for the function class $\mathcal{G}_{pol}$, the implicit definition of the bandwidth implies that $h^* \asymp T^{-\frac{1}{2+2\Delta\beta}}$. This leads to the convergence rate

$$\sup_{g\in\mathcal{G}_{\mathrm{pol}}(\beta,C_\varphi,c_\varphi,C_g,C)} \mathbb{E}_g\left[\|g-\widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})}\right] = O\left(T^{-\frac{1}{2+2\Delta\beta}}\right),$$

which is optimal for estimating g with $\mathrm{L}^2(\mathbb{R})$-loss from homogeneous observations. For the class $\mathcal{G}_{\mathrm{exp}}$, we find that $h^* \asymp (\log T/(2c_\varphi\Delta))^{-\frac{1}{\alpha}}$. The corresponding convergence rate is logarithmic, $O(\Delta^{1-2\alpha}(\log T)^{-\frac{1-2\alpha}{\alpha}})$.

(Notice that in the homogeneous case, the weights coincide for all $j = 1, ..., n$ and hence cancel in the construction of the estimator.)

(ii) For non-homogeneous observations, the rates of convergence are intimately connected to the maximal distance between the observation times. Consider the class $\mathcal{G}_{\mathrm{pol}}$. Part (i) of Proposition 3.4 then implies that $h^* < T^{-\frac{1}{2+2\beta\Delta_{max}}}$ and consequently,

$$\sup_{g\in\mathcal{G}_{\mathrm{pol}}(\beta,C_\varphi,c_\varphi,C_g,C)} \mathbb{E}_g\left[\|g-\widehat{g}_{h^*}\|^2_{\mathrm{L}^2(\mathbb{R})}\right] = O\left(T^{-\frac{1}{2+2\beta\Delta_{max}}}\right).$$

In particular, for Δ_{max} small, we approach the rate of convergence $T^{-\frac{1}{2}}$ which is optimal for estimating a density g with $\mathrm{L}^2(\mathbb{R})$-loss when $|\mathcal{F}g(u)| \sim |u|^{-1}$. The estimation of g can hence be understood in analogy with density deconvolution whenever the observations are low frequent, and in analogy with standard density estimation when Δ_{max} is small.

In the compound Poisson case, for $\rho > 0$, we find that $h^* \asymp (\log T)^{-\frac{1}{\rho}}$. The rate of convergence is $(\log T)^{\frac{1}{\rho}}T^{-1}$ and hence coincides with the optimal convergence rate for standard density estimation problems with supersmooth densities. In this particular case, Δ_{max} does not affect the rate of convergence but only appears as a constant factor.

Local regularity

It has been pointed out in the preceding section that the global regularity of g is linked to the decay behavior of $|\varphi|$, which influences the rates of convergence to be obtained. In particular, an exponential decay of $|\varphi|$ will always lead to slow logarithmic rates of convergence.
However, one may as well be interested in the estimation of g on some compact set Ω bounded away from the origin. The Lévy density and hence the function g may then be arbitrarily regular and even infinitely differentiable on Ω. In what follows, we investigate rates of convergence over nonparametric classes of locally smooth functions.

Recall that for an open interval D, the *Hölder class* $\mathcal{H}_D(a, L, R)$ consists of those functions f defined on D, whose absolute value is bounded above by R, which are $\lfloor a \rfloor$-times continuously differentiable and for which

$$\sup_{x \neq y \in D} \frac{|f^{(\lfloor a \rfloor)}(x) - f^{(\lfloor a \rfloor)}(x)|}{|x-y|^{\alpha - \lfloor a \rfloor}} \leq L.$$

holds. In the present case, $\lfloor a \rfloor$ denotes the largest integer which is smaller than (but not equal to) α.

Let us introduce nonparametric classes of locally Hölder regular functions and corresponding Lévy processes. In the sequel, $\mathcal{G}(a, D, L, R, C_1, \beta, C_2, C_3)$ denotes the class of functions g for which the following holds:

(i) There exists a Lévy process X for which (A1)-(A4) holds, with Lévy density $\eta(x) = g(x)/x$.

(ii) $g|_D$ can be extended to a function $\widetilde{g} \in \mathcal{H}_{\mathbb{R}}(a, L, R)$ and $\|g\|_{\mathrm{L}^1(\mathbb{R})} + \|\widetilde{g}\|_{\mathrm{L}^1(\mathbb{R})} < C_1$.

(iii) $\forall u \in \mathbb{R}_+ : |\varphi_{X_1}(u)| \geq (1 + C_2|u|^2)^{-\frac{\beta}{2}}$.

(iv) X has a finite fourth moment and $C_4^{\pm} \leq C_3$.

The following lemma gives a bound on the bias term for locally Hölder regular functions.

Lemma 3.3. *Let Ω be a compact interval, bounded away from the origin. Assume that there exists an open interval $D \supseteq \Omega$ such that $g|_D \in \mathcal{H}_D(a, L, R)$ and $g|_D$ can be extended to a function $\widetilde{g} \in \mathrm{L}^1(\mathbb{R}) \cap \mathcal{H}_{\mathbb{R}}(a, L, 2R)$. Let the kernel be chosen such that $\mathrm{K} \in \mathrm{L}^2(\mathbb{R})$, K has the order $a > 0$ and moreover, for some constant C_{K},*

$$\forall x \in \mathbb{R} : |\mathrm{K}(x)| \leq C_{\mathrm{K}} |x|^{-a-1}. \tag{3.3}$$

Then we can estimate for some positive constant C_b depending on the choice of K, on L, R and on Ω and D

$$\|g - \mathrm{K}_h * g\|^2_{\mathrm{L}^2(\Omega)} \leq C_b (1 + \|g\|_{\mathrm{L}^1(\mathbb{R})} + \|\widetilde{g}\|_{\mathrm{L}^1(\mathbb{R})}) h^{2a}.$$

Comment. It is important to realize that the condition (3.3) on the kernel is crucial. The sinc-kernel is thus not a reasonable choice in the present situation, for estimating g on a compact set bounded away from the origin. This is a consequence of the fact that, no matter how g may be locally, it will have discontinuity at zero when the jump activity is infinite, which implies that the function is globally non-smooth.

The following proposition is a direct consequence of the bound on the bias given above, combined with Theorem 3.1.

Proposition 3.4. *Let K be such that the assumptions summarized in Lemma 3.3 are met. Moreover, let K be supported on $[-1, 1]$. Let h^* be implicitly defined as the solution of*

$$1 = \sum_{j=1}^{n} \Delta_j h^{2\beta\Delta_j + 2a + 1}. \tag{3.5}$$

Then

$$\sup_{g \in \mathcal{G}(a, D, L, R, C_1, \beta, C_2, C_3)} \mathbb{E}_g \left[\|g - \widehat{g}_{h^*}\|^2_{\Omega} \right] = O\left(h^{*2a} \right).$$

Example. Again, we investigate how Δ_{max} influences the rate of convergence to be obtained. We find that $h^* \lesssim h'$, with

$$h' = T^{-\frac{1}{2a+2\Delta_{max}\beta+1}}.$$

The resulting rate of convergence is faster than or equal to $O(T^{-\frac{2a}{2a+2\Delta_{max}\beta+1}})$. This implies that rate approaches the rate of convergence $T^{-\frac{2a}{2a+1}}$ which is known the be optimal for estimating g (locally) with L^2-loss when continuous time observations of the process are available.

Comment. In the same spirit, we may consider the case where $|\varphi(u)|$ decays exponentially, $|\varphi(u)| \sim e^{-c|u|^\gamma}$. It can then be shown that, with optimal choice of h^*, one attains convergence of order $(\log T)^{-\frac{2a}{\gamma}}$.

3.3 Lower bounds

In the sequel, we show that the rates of convergence presented in the preceding paragraphs are minimax optimal. However, to avoid some technical difficulties, we content ourselves with considering the case of polynomially decaying characteristic functions.

Theorem 3.5. *(i) Let h^* be defined according to (3.2). Then there exists a positive constant c such that*

$$\liminf_{n\to\infty} \inf_{\breve{g}} \sup_{g\in\mathcal{G}_{pol}(\beta,C_\varphi,c_\varphi,C_g,C)} \mathbb{E}_g\left[\|g-\breve{g}\|^2_{\mathrm{L}^2(\mathbb{R})}\right](h^*)^{-1} \geq c.$$

The infimum is taken over the collection of estimators $\widetilde{g}_n$ based on the observations $X_{t_1},\cdots,X_{t_n}$.

(ii) Let h^* be defined according to (3.4). Then there exists a positive constant c such that

$$\liminf_{n\to\infty} \inf_{\breve{g}} \sup_{g\in\mathcal{G}(a,D,L,R,C_1,\beta,C_2,C_3)} \mathbb{E}_g\left[\|g-\breve{g}\|^2_{\mathrm{L}^2(\Omega)}\right](h^*)^{-2a} \geq c.$$

4 Estimating the distributional density

Let us have a look at the situation where one is interested in estimating the density f of X_1 rather than the underlying Lévy measure. We can now drop the technical assumptions (A1)-(A4) on the process, thus allowing a non-zero Gaussian part, a high activity of small jumps and the existence of a drift. In the sequel, it is only assumed that the density f of X_1 exists and is square integrable on the whole real axes, thus excluding processes of compound Poisson type.
Without specifying any particular assumptions on the Lévy measure, drift or Gaussian part, the estimation procedure proposed in Section 3 can still be used to build an estimator of the derivative of the characteristic exponent,

$$\widehat{\Psi'}(z) := \frac{\widehat{p}(z)}{\widetilde{q}(z)}.$$

Since $\Psi(0)=0$ holds by definition of the characteristic exponent, a corresponding estimator of the characteristic exponent can be defined as follows.

$$\widehat{\Psi}(u) := \int_0^u \widehat{\Psi}'(z)\,\mathrm{d}z = \int_0^u \frac{\widehat{p}(z)}{\widetilde{q}(z)}\,\mathrm{d}z.$$

The characteristic function $\varphi(u)=\exp(\Psi(u))$ of X_1 can then be estimated by $\check{\varphi}(u) := e^{\widehat{\Psi}(u)}$. However, there is a priori no guarantee that $\check{\varphi}$ is a characteristic function and the absolute value may be larger than one. For this reason, we introduce an additional threshold, thus defining the final estimator of φ

$$\widehat{\varphi}(u) := \frac{\check{\varphi}(u)}{\max\{1, |\check{\varphi}(u)|\}}.$$

Given a kernel function K and bandwidth $h>0$, the density f is then estimated using kernel smoothing and Fourier inversion,

$$\widehat{f}_h(x) := \frac{1}{2\pi}\int e^{-iux}\mathcal{F}\mathrm{K}(hu)\widehat{\varphi}(u)\,\mathrm{d}u.$$

Theorem 4.1. *Let* K *be supported on* $[-1,1]$*. Assume that* $\mathbb{E}\left[|X|^{4m}\right]<\infty$*.*
Then there exists some $C:\mathbb{R}^2\to\mathbb{R}_+$ *which is monotonously increasing in both components such that*

$$\mathbb{E}\big[\|f-\widehat{f}_h\|_{\mathrm{L}^2}^2\big] \le 2\|f-\mathrm{K}_h * f\|_{\mathrm{L}^2}^2 + C(\Delta_{Max}, C_4^{\pm})\int_{-1/h}^{1/h}|\varphi(u)|^2 C_{\Psi,1}(u)\int_0^u \frac{1}{|q(x)|}\,\mathrm{d}x\,\mathrm{d}u$$
$$+\int_{-1/h}^{1/h}\Big(C_{\Psi,1}(u)\int_0^u\frac{1}{|q(x)|}\,\mathrm{d}x\Big)^m \max\Big\{1, C_{\Psi,2}(u)\int_0^u \frac{1}{|q(x)|}\,\mathrm{d}x\Big\}^m\,\mathrm{d}u,$$

with

$$C_{\Psi,1}(u) := (\|\Psi'\|_{\mathrm{L}^2,u}^2 + \|\Psi''\|_{\mathrm{L}^1,u})\vee 1 \quad \textit{and} \quad C_{\Psi,2}(u) := \|\Psi'\|_{\infty,u}.$$

Rates of convergence

For $\beta>\frac{1}{2}$, let $\mathcal{F}_{\mathrm{pol}}(\beta,m,C_1,C_2,C_3,C)$ be the class of densities corresponding to an infinitely divisible distribution for which the following holds.

(i) For the characteristic function φ of f,

$$\forall u\in\mathbb{R}: (1+C_1|u|)^{-\beta} \le |\varphi(u)| \le (1+C_2|u|)^{-\beta}.$$

(ii) Ψ' is square integrable and Ψ'' is integrable, with

$$\|\Psi'\|_{\mathrm{L}^2(\mathbb{R})}^2 + \|\Psi''\|_{\mathrm{L}^1(\mathbb{R})} + \|\Psi'\|_\infty < C_3$$

(iii) The random variable X_1 with density f has finite moments up to order $4m$ and $C_{4m}^{\pm}\le C$.

This function class contains, for example, the densities of gamma- or bilateral gamma distributions.

For $\alpha \in (0,2]$, let $\mathcal{F}_{\mathrm{exp}}(\alpha, m, C_1, C_2, C_3, C)$ be the class of infinitely divisible densities such that

(i) $\forall u \in \mathbb{R} : C_1 \exp(-c|u|^\alpha) \leq |\varphi(u)| \leq C_2 \exp(-c|u|^\alpha)$

(ii) $\|\Psi'\|^2_{\mathrm{L}^2,u} + \|\Psi'\|_{\mathrm{L}^1,u} + \|\Psi'\|_{\infty,u} \leq \begin{cases} C_3, & \text{for } \alpha < \frac{1}{2} \\ C_3(\ln|u| \vee 1), & \text{for } \alpha = \frac{1}{2} \\ C_3|u|^{2\alpha-1}. & \text{for } \alpha > \frac{1}{2}. \end{cases}$

(iii) The moments up to order $4m$ are finite and $C^{\pm}_{4m} \leq C$.

Typical examples are tempered stable laws with index of stability α, or (for $\alpha = 2$) processes with non-zero Gaussian part.

In the sequel, we define, for the exponential or polynomial decay scenario, respectively,

$$|\varphi_{\mathrm{reg}}(u)| := (1 + |u|)^{-\beta} \quad \text{or} \quad |\varphi_{\mathrm{reg}}(u)| := \exp(-c|u|^\alpha).$$

Moreover,

$$q_{\mathrm{reg}}(u) := \sum_{j=1}^{n} \Delta_j |\varphi_{\mathrm{reg}}(u)|.$$

The convergence rates over nonparametric function classes summarized in the following Proposition are immediate consequences of Theorem 4.1, performing the compromise between the bias and variance term.

Proposition 4.2. *Let* K *be the sinc-kernel,* $\mathcal{F}\mathrm{K} = \mathbb{1}_{[-1,1]}$.

(i) Consider $\beta > \frac{1}{2}$ *and* $\alpha \in (0,1)$*. Then, with* h^* *implicitly defined implicitly, as the solution to*

$$|\varphi_{reg}(1/h)|^2 = \Big(\int_0^{1/h} \frac{1}{q_{reg}(x)}\, \mathrm{d}x \Big)^k, \tag{4.1}$$

we derive that

$$\sup_{f \in \mathcal{F}_{pol}(\beta, m, C_1, C_2, C_3, C)} \mathbb{E}\big[\|f - \widehat{f}\|^2_{\mathrm{L}^2(\mathbb{R})}\big] = O\big((h^*)^{2\beta-1} \vee T^{-1}\big)$$

and

$$\sup_{f \in \mathcal{F}_{exp}(\alpha, m, C_1, C_2, C_3, C)} \mathbb{E}\big[\|f - \widehat{f}\|^2_{\mathrm{L}^2(\mathbb{R})}\big] = O\big(h^{\alpha-1} \exp(-2c(1/h^*)^\alpha) \vee T^{-1}\big).$$

(ii) Consider $\alpha = \frac{1}{2}$ *or* $\alpha \in (\frac{1}{2}, 2]$*. Let* h^* *be the solution to*

$$|\varphi_{reg}(1/h)|^2 = \Big(\ln(1/h) \int_0^{1/h} \frac{1}{q_{reg}(x)}\, \mathrm{d}x \Big)^k \quad \text{or} \quad |\varphi_{reg}(1/h)|^2 = \Big((1/h)^{2\alpha-1} \int_0^{1/h} \frac{1}{q_{reg}(x)}\, \mathrm{d}x \Big)^k.$$

Then it follows that

$$\sup_{f\in\mathcal{F}_{exp}(\alpha,m,C_1,C_2,C_3,C)} \mathbb{E}\big[\|f-\widehat{f}\|^2_{\mathrm{L}^2(\mathbb{R})}\big] = O\big((h^{\alpha-1}\vee 1)\exp(-2c(1/h^*)^\alpha)\vee T^{-1}\big).$$

Again, the implicit definition of the smoothing parameter is not particularly intuitive so we have a look at some examples to understand how the convergence rates are connected to the maximal distance between the observation times.

Examples.

(i) In the polynomial decay scenario, (4.2) implies that $h^* \gtrsim h'$, with

$$h' = T^{-\frac{1}{2\Delta_{max}\beta+1+\frac{2\beta}{k}}},$$

and the corresponding convergence rate is faster than or equal to $O\Big(T^{-\frac{2\beta-1}{2\Delta_{max}\beta+1+\frac{2\beta}{k}}}\vee T^{-1}\Big)$. This implies, in particular, that for $\beta > 1$, without any further restriction on the regularity of the observation times, the estimator attains the parametric rate of convergence if a finite $4k$-th moment exists and $\Delta_{max}\leq 1-\frac{1}{\beta}-\frac{1}{k}$.

(ii) In the exponential decay scenario with $\alpha < \frac{1}{2}$, (4.2) implies that $h^* \gtrsim h'$, with

$$h' = \Big(\frac{\ln T - \ln(\ln T)^{\frac{1-\alpha}{\alpha}}}{2c(\Delta_{max}+\frac{1}{k})}\Big)^{\frac{1}{\alpha}}.$$

The corresponding rate of convergence is faster than or equal to $(\ln T)^\gamma T^{-\frac{1}{\Delta_{max}+\frac{1}{k}}}\vee T^{-1})$, with $\gamma = \frac{(1-\alpha)(\Delta_{max}1+1/k}{(\Delta_{max}+1/k)\alpha}$.

It follows that the parametric rate is attained whenever the $4m$-th moment is finite and $\Delta_{max} < 1-\frac{1}{k}$.

For exponential decay with $\alpha\geq 1/2$, the arguments are the same, apart from an additional logarithmic loss.

5 Numerical examples

5.1 Practical calculation of the weights and data driven bandwidth selection

The weights assigned to each of the $\widehat{\varphi}_j$ are crucial for the construction of the estimator and the ideal weights $w_j^* = \overline{\varphi_j}$ are not feasible to actually compute.
On may consider the following selection algorithm for the weights. The interval $[0,\Delta_{max}]$ is divided into K_n disjoint intervals of equal length. Set $\mathcal{J}_k = \{\ell : \Delta_\ell \in I_k\}$. Then, for $\Delta_j\in I_k$, a biased estimator of $w_j = \overline{\varphi}_j$ can be constructed, setting

$$\widehat{w}_j := \frac{1}{|\mathcal{J}_k|}\sum_{\ell\in\mathcal{J}_k} e^{-iuZ_\ell}.$$

It can be shown that this estimation procedure has good theoretical properties and it preserves, up to some logarithmic loss, the upper risk bound and convergence rates which have been derived for the oracle estimator when the number K_n of sub-intervals is logarithmic in n. However, in numerical examples, the practical performance of the estimator turns out to be unsatisfactory. For this reason, we cannot recommend this procedure in applications.
Instead, we propose an iterative selection algorithm for the weights. A preliminary estimator $\widehat{\Psi}$ of Ψ is calculated, using the initial weights $w_j \equiv 1$. Improved estimators of the weights are then given by $\widehat{w}_j = \overline{\exp(\Delta_j \widehat{\Psi})}$. These improved estimators are then applied to build a new estimator of Ψ. The procedure is iterated until the $\mathrm{L}^2([-\sqrt{T}, \sqrt{T}])$-distance between the empirical weights is sufficiently small.
The algorithm can be summarized as follows.

$\forall j : \widehat{w}_{j,0} \leftarrow 0$
$\forall j : \widehat{w}_{j,1} \leftarrow 1$
$m \leftarrow 1$
while$(\exists j : \|\widehat{w}_{j,m-1} - \widehat{w}_{j,m}\|^2_{\mathrm{L}^2([-\sqrt{T},\sqrt{T}])} > 1/T)$
$\{$
$\forall j : w_j \leftarrow \widehat{w}_{j,m}$
$\widehat{\Psi}(u) \leftarrow \int_0^u \widehat{p}(z)/\widetilde{q}(z)\, \mathrm{d}z$
$\forall j : \widehat{w}_{j,m+1} \leftarrow \overline{\exp(\Delta_j \widehat{\Psi})}$
$m \leftarrow m + 1$
$\}$
$return(w_1, \cdots, w_n)$

5.2 Estimation of the jump measure

We consider the particular case where the kernel is specified to be the sinc-kernel and g is estimated with $\mathrm{L}^2(\mathbb{R})$-loss. We consider the collection of cutoff parameters $\mathcal{M} = \{m \in \mathbb{N} : m \leq \sqrt{T}\}$. The data driven cutoff parameter $\widehat{m}$ is calculated, using a leave-p-out cross validation strategy: Given any subset $P = \{n_1, .., n_p\} \subseteq \{1, \cdots, n\}$ with p elements, we calculate the estimator $(\widehat{\Psi}')^{(P)}$ of Ψ', based on the data set $\{Z_j,\ j \in P\}$ as well as the estimator $(\widehat{\Psi}')^{(-P)}$ based on $\{Z_j,\ j \notin P\}$. In the definition of $(\widehat{\Psi}')^{(P)}$ we omit the regularization of $\widehat{q}$ in the denominator.
By the oracle cutoff parameter m^*, we understand the minimizer of the L^2-loss,

$$m^* = \operatorname*{argmin}_{m \in \mathcal{M}} \int |\widehat{g}_m(x) - \widehat{g}(x)|^2 \, \mathrm{d}x.$$

By the Plancherel formula,

$$m^* = \operatorname*{argmin}_{m \in \mathcal{M}} \int_{-m}^{m} |\widehat{\Psi}'(u)|^2 \, \mathrm{d}u - 2 \operatorname{Re} \int_{-m}^{m} \widehat{\Psi}'(u) \overline{\Psi'(u)} \, \mathrm{d}u =: \ell(g, \widehat{g}_m).$$

Since this true loss function is not feasible to compute, $\ell(g, \widehat{g})$ is approximated by the empirically accessible quantity

$$\widehat{\ell}(g, \widehat{g}_m) := \int_{-m}^{m} |\widehat{\Psi}'(u)|^2 \, \mathrm{d}u - 2 \binom{n}{p}^{-1} \sum_{P \subseteq \{1, \ldots, n\}} \operatorname{Re} \int_{-m}^{m} (\widehat{\Psi}')^{(P)}(u) \overline{(\widehat{\Psi}')^{(\text{-}P)}(u)} \, \mathrm{d}u,$$

and the data driven cutoff parameter is defined to be

$$\widehat{m} = \inf_{m \in \mathcal{M}} \widehat{\ell}(g, \widehat{g}_m).$$

We work with the choice $p = n/10$. The constant threshold value κ is set equal to 1. However, when considering all subsets of size p, the numerical complexity of the algorithm explodes. In praxis, we do not consider all subsets, but content ourselves with splitting the data into $n/100$ disjoint blocks $B_1, \cdots, B_{n/100}$ of equal size and calculate the leave-p-out estimators based on all subsets of the form $P = B_j \cup B_{j+1} \cup \ldots \cup B_{j+9},\ j = 1, \cdots, n/10 - 9$.
In the definition of the estimator, the distances between the observations may be fairly arbitrary and no particular structural assumptions are imposed, despite of the fact that the distances are bounded from above. We calculate, in a preliminary step, distances between the observations which are independent realizations of a $\mathcal{U}[u,t]$-distribution. We consider the cases where $t = 6$ or $t = 2$.
The cross validated estimator is compared to the "estimator" with oracle choice of the weights and oracle cutoff of the bandwidth. Moreover, in order to illustrate the relevance of the proper choice of the weights, we give, for comparison, the "estimator" where all weights are set equal to 1. Again, we use in this case the oracle choice of the bandwidth.
The following examples are being considered.

(i) X is a Gamma-process with parameters 3 and 2, so $X_1 \sim \Gamma(3,2)$.

(ii) X is a compound Poissson process with intensity parameter 3 and standard normally distributed jumps.

(iii) X is a symmetric bilateral Gamma process with parameters 2 and 4, so X_1 is the difference between two independent $\Gamma(2,4)$-distributed random variables.

Based on 500 independent realizations, we calculate the empirical risk r_{or} of the "estimator" with oracle weights and oracle bandwidth, the empirical risk r_{ad} of the estimator with adaptive bandwidth selection and data driven choice of the weights, as well as the empirical risk r_{eq} corresponding the the procedure with equal weights and oracle bandwidth.
In Table 1 below, we summarize the result, for observation distances which are generated using a $\mathcal{U}([0,6])$-distribution. In Table 2, the results are given when the observation distances are generated, using a $\mathcal{U}([0,2])$-distribution.

5.2.1 Density estimation

In the same way as in the preceding section, we define the estimators $(\widehat{\Psi'})^{(P)}$ and $(\widehat{\Psi'})^{(-P)}$. The corresponding estimators of the characteristic function are $\widehat{\varphi}^{(P)}(u) := \exp(\int_0^u (\widehat{\Psi'})^{(P)}(x)\,\mathrm{d}x)$ and $\widehat{\varphi}^{(-P)}(u) := \exp(\int_0^u (\widehat{\Psi'})^{(-P)}(x)\,\mathrm{d}x)$. The cutoff parameter is then defined to be

$$\widehat{m} := \inf_{m \in \mathcal{M}} \int_{-m}^{m} |\widehat{\varphi}(u)|^2\,\mathrm{d}u - 2\binom{n}{p}^{-1} \sum_{P \subseteq \{1,\ldots,n\}} \int_{-m}^{m} \operatorname{Re}\left(\widehat{\varphi}^{(P)}(u)\overline{\widehat{\varphi}^{(-P)}(u)}\right)\mathrm{d}u.$$

Again, we specify $p = n/10$ and consider, in praxis, only a part of the subsets of size p.
The cross-validated estimator with empirical weights is once more compared to the estimator with oracle choice of the weights and of the cutoff parameter, as well as to the estimator where all weights are identical 1.
We consider the following examples:

(i) Gamma-process with parameters 3 and 2.

(ii) Bilateral gamma-process with parameters 2 and 4.

(iii) Brownian motion with variance 1 and drift 2.

	$\Gamma(3,2)$			$\text{Cpois}(\mathcal{N}(0,1),3)$		
n	r_{or}	r_{ad}	r_{eq}	r_{or}	r_{ad}	r_{eq}
1000	0.5152	1.1669	1.3790	0.1663	0.3330	0.7350
5000	0.2936	0.5350	1.2297	0.0495	0.0981	0.4845
10000	0.2097	0.3456	1.1378	0.0268	0.0529	0.3828

	$b\Gamma(2,4,2,4)$		
n	r_{or}	r_{ad}	r_{eq}
1000	0.2845777	0.4068848	0.6049949
5000	0.1123	0.1533	0.5230
10000	0.0687	0.0871	0.4719

Table 1: $\mathcal{U}([0,6])$-distributed observation distances

	$\Gamma(3,2)$			$\text{Cpois}(\mathcal{N}(0,1),3)$		
n	r_{or}	r_{ad}	r_{eq}	r_{or}	r_{ad}	r_{eq}
1000	0.3483	0.6648	0.8352	0.0784	0.1975	0.1937
5000	0.1632	0.2374	0.6637	0.0201	0.0388	0.0596
10000	0.1104	0.14625	0.5786	0.0111	0.0189	0.0326

	$b\Gamma(2,4,2,4)$		
n	r_{or}	r_{ad}	r_{eq}
1000	0.1682	0.2484	0.3617
5000	0.0460	0.0574	0.2314
10000	0.0246	0.0269	0.1817

Table 2: $\mathcal{U}([0,2])$-distributed observation distances

The results are summarized in Table 3 and Table 4 below.
Discussion: It is surprising to realize that for medium to large sample sizes, there is no visible difference in the performance of the oracle estimator and the estimator with data driven choice of the bandwidth and the weights, whereas the procedure with all weights set equal to 1 performs substantially worse.

	$\Gamma(3,2)$			$b\Gamma(2,4,2,4)$		
n	r_{or}	r_{ad}	r_{eq}	r_{or}	r_{ad}	r_{eq}
1000	0.00319	0.00418	0.06389	0.00369	0.00460	0.01777
5000	0.00095	0.00099	0.03510	0.00083	0.00088	0.00760
10000	0.00060	0.00061	0.02889	0.00041	0.00041	0.00501

	$\mathcal{N}(2,1)$		
n	r_{or}	r_{ad}	r_{eq}
1000	0.00140	0.00288	0.03632
5000	0.00065	0.00065	0.01972
10000	0.00055	0.00056	0.01547

Table 3: Estimation of the distributional density, with $\mathcal{U}([0,6])$-distributed observation distances

	$\Gamma(3,2)$			$b\Gamma(2,4,2,4)$		
n	r_{or}	r_{ad}	r_{eq}	r_{or}	r_{ad}	r_{eq}
1000	0.00168	0.00178	0.00786	0.00185	0.00195	0.00305
5000	0.00057	0.00057	0.00205	0.00035	0.00036	0.00074
10000	0.00043	0.00044	0.00129	0.00018	0.00018	0.00040

	$\mathcal{N}(2,1)$		
n	r_{or}	r_{ad}	r_{eq}
1000	0.00089	0.00090	0.00419
5000	0.00056	0.00057	0.00217
10000	0.00052	0.00052	0.01533

Table 4: Estimation of the distributional density, with $\mathcal{U}([0,2])$-distributed observation distances

6 Proofs

6.1 Preliminaries

Lemma 6.1. *For arbitrary $m \in \mathbb{N}$, there exists a positive constant C depending on m such that*

$$\mathbb{E}\big[\,|\widehat{q}(u)-q(u)|^{2m}\,\big] \le C\big(\sum_{j=1}^{n}\Delta_j^2|w_j(u)|^2\big)^m = C\sigma(u)^{2m}.$$

Proof. By the Rosenthal inequality (see, e.g. Ibragimov and Sharakhmetov (2002)) and the fact that $|\widehat{\varphi}_j-\varphi_j|$ is bounded by 2, there exist constants C' and C depending on m such that

$$\begin{aligned}
&\mathbb{E}\big[\,|\widehat{q}(u)-q(u)|^{2m}\,\big]\\
\le& C'\max\left\{\left(\sum_{j=1}^{n}\Delta_j^2|w_j|^2\mathbb{E}\big[|\widehat{\varphi}_j(u)-\varphi_j(u)|^2\big]\right)^m, \sum_{j=1}^{n}\Delta_j^{2m}|w_j|^{2m}\mathbb{E}\big[|\widehat{\varphi}_j(u)-\varphi_j(u)|^{2m}\big]\right\}\\
\le& C\left(\sum_{j=1}^{n}\Delta_j^2|w_j|^2\right)^m = C\sigma(u)^{2m}.
\end{aligned}$$

This is the desired result. ◻

Lemma 6.2. *For arbitrary $m\in\mathbb{N}$, there exists a constant C depending on κ, Δ_{max} and m such that*

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^m\right] \le C\min\left\{\frac{1}{|q(u)|^m}, \frac{1}{|q(u)|^{\frac{3m}{2}}}\right\}$$

Proof. Consider first the case where $|q(u)|\le 2\sigma(u)$. Using the fact that, by definition of $1/\widetilde{q}$ and the assumption on q,

$$\frac{1}{|\widetilde{q}(u)|}\le\frac{1}{\sigma(u)}<\frac{2}{|q(u)|},$$

as well as the estimate

$$\begin{aligned}
\sigma(u) =& \left(\sum_{j=1}^{n}\Delta_j^2|w_j(u)|^2\right)^{\frac12} \le \left(\Delta_{max}\sum_{j=1}^{n}\Delta_j|w_j(u)|^2\right)^{\frac12} = \left(\Delta_{max}\sum_{j=1}^{n}\Delta_j|\varphi_j(u)|^2\right)^{\frac12}\\
=&\Delta_{max}^{\frac12}q(u)^{\frac12},
\end{aligned}\tag{6.1}$$

we find that

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^m\right] \le 3^m\frac{1}{|q(u)|^m}\le 6^m\frac{\sigma(u)^m}{|q(u)|^{2m}}\le\frac{(6\sqrt{\Delta_{max}})^m}{|q(u)|^{\frac{3m}{2}}}.$$

Next, assume that $|q(u)|>2\sigma(u)$ and, in addition, $\sigma(u)\ge\kappa$. We may then use the decomposition

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^m\right] = \frac{1}{|q(u)|^m}\,\mathbb{P}\left(\{|\widehat{q}(u)|<\sigma(u)\}\right)+\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widehat{q}(u)}\right|^m\mathbb{1}_{\{|\widehat{q}(u)|\ge\sigma(u)\}}\right].$$

By the Markov inequality,

$$\mathbb{P}\left(\left\{|\widehat{q}(u)|<\sigma(u)\right\}\right) \leq \mathbb{P}\left(\left\{|q(u)-\widehat{q}(u)|>\frac{|q(u)|}{2}\right\}\right) \leq 2^m \frac{\mathbb{E}\left[[]\,|q(u)-\widehat{q}(u)|^m\right]}{|q(u)|^m}$$
$$\leq 2^m \frac{\sigma(u)^m}{|q(u)|^m} \leq (2\sqrt{\Delta_{max}})^m \frac{1}{|q(u)|^{\frac{m}{2}}}.$$

On the other hand, using Lemma 6.1 and formula () we can estimate for a constant C depending on m,

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widehat{q}(u)}\right|^m \mathbb{1}_{\{|\widehat{q}(u)|\geq\sigma(u)\}}\right] \leq 2^m\Big(\frac{\mathbb{E}\left[|q(u)-\widehat{q}(u)|^m\right]}{|q(u)|^{2m}}+\frac{\mathbb{E}\left[|q(u)-\widehat{q}(u)|^{2m}\right]}{|q(u)|^{2m}\sigma(u)^m}\Big)$$
$$\leq C\frac{\sigma(u)^m}{|q(u)|^{2m}} \leq C\min\left\{\frac{1}{|q(u)|^m},\frac{\sigma(u)^m}{|q(u)|^{2m}}\right\} \leq C\Big\{\frac{1}{|q(u)|^m},\frac{\sqrt{\Delta_{max}}^m}{|q(u)|^{\frac{3m}{2}}}\Big\}. \tag{6.2}$$

Finally, consider the case where $|q(u)|>2\sigma(u)$ and $\sigma(u)<\kappa$. We can then decompose

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^m\right]=\frac{1}{|q(u)|^m}\mathbb{P}\left(\left\{|\widehat{q}(u)|<\kappa\right\}\right)+\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widehat{q}(u)}\right|^m\mathbb{1}_{\{|\widehat{q}(u)|\geq\kappa\}}\right]. \tag{6.3}$$

We distinguish between two different sub-cases.
a) In addition to $|q(u)|>2\sigma(u)$ and $\sigma(u)<\kappa$, $|q(u)|\leq 2\kappa$. We may then estimate

$$\frac{1}{|q(u)|^m}=(2\kappa)^{\frac{m}{2}}\frac{1}{|q(u)|^m(2\kappa)^{\frac{m}{2}}} \leq (2\kappa)^{\frac{m}{2}}\frac{1}{|q(u)|^{\frac{3m}{2}}}$$

and, moreover,

$$\left|\frac{1}{q(u)}-\frac{1}{\widehat{q}(u)}\right|^m\mathbb{1}_{\{|\widehat{q}(u)|\geq\kappa\}} \leq \frac{3^m}{|q(u)|^m} \leq \frac{(3\sqrt{2\kappa})^m}{|q(u)|^{\frac{3m}{2}}}.$$

Along with formula (), this yields the desired result for the sub-case.
b) $|q(u)|>2\sigma(u)$, $\sigma(u)<\kappa$ and, in addition, $|q(u)|>2\kappa$. In this case, another application of the Markov inequality and Lemma 6.1 gives for a constant C depending on m,

$$\mathbb{P}\left(\left\{|\widehat{q}(u)|<\kappa\right\}\right) \leq \mathbb{P}\left(\left\{|q(u)-\widehat{q}(u)|>\frac{|q(u)|}{2}\right\}\right) \leq C\frac{\sigma(u)^m}{|q(u)|^m} \leq C\sqrt{\Delta_{max}}\frac{1}{|q(u)|^{\frac{3m}{2}}}.$$

Next, arguing along the same lines as in formula (), we find that

$$\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widehat{q}(u)}\right|^m\mathbb{1}_{\{|\widehat{q}(u)|\geq\kappa\}}\right] \leq C\Big(\frac{\sigma(u)^m}{|q(u)|^{2m}}+\frac{\sigma(u)^{2m}}{|q(u)|^{2m}\kappa^m}\Big) \leq C\frac{\sigma(u)^m}{|q(u)|^{2m}}$$
$$\leq C\min\left\{\frac{2^m}{|q(u)|^m},\frac{2^{m+1}\Delta_{max}^{\frac{m}{2}}}{|q(u)|^{\frac{3m}{2}}}\right\}.$$

This completes the proof. $\square$

Recall that X_+ and X_- are independent Lévy processes without drift and with jump measures $\nu_+(\,\mathrm{d}x)=\nu|_{(0,\infty)}(\,\mathrm{d}x)$ and $\nu_-(\,\mathrm{d}x)=\nu|_{(-\infty,0)}(\,\mathrm{d}(\text{-}x))$

Lemma 6.3. *For arbitrary $\Delta \geq 0$ and $m \in \mathbb{N}$,*

$$\mathbb{E}\left[|X_\Delta|^m\right] \leq \max\{\Delta, \Delta^m\}\big(\mathbb{E}\left[(X_+)_1^m\right] + \mathbb{E}\left[(X_-)_1^m\right]\big) = \max\{\Delta, \Delta^m\}C_m^\pm$$

Proof. Since X_Δ has the same distribution as $(X_+)_\Delta - (X_-)_\Delta$, we have

$$\mathbb{E}\left[|X_\Delta|^m\right] = \mathbb{E}\left[|(X_+)_\Delta - (X_-)_\Delta|^m\right] \leq \mathbb{E}\left[\max\{(X_+)_\Delta^m, (X_-)_\Delta^m\}\right] \leq \mathbb{E}\left[(X_+)_\Delta^m\right] + \mathbb{E}\left[(X_-)_\Delta^m\right]. \quad (6.4)$$

Let $\mathcal{A} := \{\alpha \in \{1, ..., m\}^m : \sum_{\ell=1}^m \ell\alpha_\ell = m\}$. By Faà di Bruno's theorem, there exist universal constants $c_\alpha, \alpha \in \mathcal{A}$ such that

$$\mathbb{E}\left[(X_+)_\Delta^m\right] = i^{-m}\varphi_{(X_+)_\Delta}^{(m)}(0) = i^{-m}\sum_{\alpha\in\mathcal{A}} c_\alpha \prod_{\ell=1}^m \left(\Delta\Psi^{(\ell)}(0)\right)^{\alpha_\ell} = \sum_{\alpha\in\mathcal{A}} c_\alpha \prod_{\ell=1}^m \Big(\Delta\int x^\ell \nu_+(\,\mathrm{d}x)\Big)^{\alpha_\ell},.$$

Since ν_+ is supported on the positive half axes, this implies

$$\mathbb{E}\left[(X_+)_\Delta^m\right] \leq \max\{\Delta, \Delta^m\}\sum_{\alpha\in\mathcal{A}} c_\alpha \prod_{\ell=1}^m \Big(\int x^\ell \nu_+(\,\mathrm{d}x)\Big)^{\alpha_\ell} = \max\{\Delta, \Delta^m\}\mathbb{E}\left[(X_+)_1^m\right].$$

Arguing along the same lines gives $\mathbb{E}\left[(X_-)_\Delta^m\right] \leq \max\{\Delta, \Delta^m\}\mathbb{E}\left[(X_-)_1^m\right]$. Combining this with formula (), gives the statement of the Lemma. $\square$

Lemma 6.4. *For any $m \in \mathbb{N}$, there exists a constant C depending on m, Δ_{Max} and $C_{2m}^\pm$ such that*

$$\mathbb{E}\Big[\big|\widehat{p}_n(u) - p(u)\big|^{2m}\Big] \leq C\max\Big\{|q(u)|^m, |q(u)|\Big\}.$$

Proof. Another application of the Rosenthal inequality, along with the fact that $|w_j| \leq 1$ and Lemma 6.3 gives, for constants C_k, $k = 1, ..., 3$ depending only on m, and some C depending on m, Δ_{Max} and $C_{2m}^\pm$,

$$\begin{aligned}
&\mathbb{E}\Big[\big|\widehat{p}(u) - p(u)\big|^{2m}\Big]\\
\leq &C_1 \max\Big\{\Big(\sum_{j=1}^n |w_j(u)|^2\mathbb{E}\left[|\widehat{\varphi}_j'(u) - \varphi_j'(u)|^2\right]\Big)^m, \sum_{j=1}^n |w_j(u)|^{2m}\mathbb{E}\left[|\widehat{\varphi}_j'(u) - \varphi_j'(u)|^{2m}\right]\Big\}\\
\leq &C_2 \max\Big\{\Big(\sum_{j=1}^n |w_j(u)|^2\mathbb{E}\left[|Z_j|^2\right]\Big)^m, \sum_{j=1}^n |w_j(u)|^{2m}\mathbb{E}\left[|Z_j|^{2m}\right]\Big\}\\
\leq &C_3 C_{2m}^\pm \max\Big\{\Big(\sum_{j=1}^n |w_j(u)|^2 \max\{\Delta_j, \Delta_j^2\}\Big)^m, \sum_{j=1}^n |w_j(u)|^{2m} \max\{\Delta_j, \Delta_j^{2m}\}\Big\}\\
\leq &C_4 C_{2m}^\pm \Delta_{\text{Max}}^{2m-1} \max\Big\{\Big(\sum_{j=1}^n \Delta_j|w_j(u)|^2\Big)^m, \sum_{j=1}^n \Delta_j|w_j(u)|^2\Big\} = C\max\left\{|q(u)|^m, |q(u)|\right\}.
\end{aligned}$$

This is the desired result $\square$

6.2 Estimation of the jump dynamics

6.2.1 Non-asymptotic upper bound

Proof of Theorem 3.1 We can estimate

$$\mathbb{E}\left[\|g-\widehat{g}_{h,n}\|_{\mathrm{L}^2}^2\right] \le 2\|g-\mathrm{K}_h * g\|_{\mathrm{L}^2(\Omega)} + 2\mathbb{E}\left[\|\mathrm{K}_h * g-\widehat{g}_{h,n}\|_{\mathrm{L}^2}^2\right].$$

The Plancherel formula gives

$$\mathbb{E}\left[\|\mathrm{K}_h * g-\widehat{g}_{h,n}\|_{\mathrm{L}^2}^2\right] = \frac{1}{2\pi}\int |\mathcal{F}\mathrm{K}_h(u)|^2\mathbb{E}\left[\left|\frac{\widehat{p}(u)}{\widetilde{q}(u)}-\frac{p(u)}{q(u)}\right|^2\right]\mathrm{d}u.$$

Now, we have the inequality

$$\left|\frac{\widehat{p}(u)}{\widetilde{q}(u)}-\frac{p(u)}{q(u)}\right|^2 \le 2\left(\frac{|\widehat{p}(u)-p(u)|^2}{|\widetilde{q}(u)|^2}+|p(u)|^2\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^2\right).$$

Consider the first summand on the right hand side of formula (6.2). We start by considering $|q(u)|<1$. Then, using the fact that $|\widetilde{q}|$ is bounded from below by κ as well as Lemma 6.4, we find that

$$\mathbb{E}\left[\frac{|\widehat{p}(u)-p(u)|^2}{|\widetilde{q}(u)|^2}\right] \le \frac{\mathbb{E}\left[|\widehat{p}(u)-p(u)|^2\right]}{\kappa^2|q(u)|^2} \le C\frac{1}{|q(u)|},$$

with some C depending on $\Delta_{max}, C_2^{\pm}$ and κ.
Next, consider the case where $|q(u)|>1$. With the Cauchy-Schwarz inequality, Lemma 6.4 and Lemma 6.2, for some C depending on Δ_{Max} , $C_4^{\pm}$ and κ,

$$\mathbb{E}\left[\frac{|\widehat{p}_n(u)-p(u)|^2}{|\widetilde{q}_n(u)|^2}\right] \le \mathbb{E}^{\frac{1}{2}}\left[|\widehat{p}_n(u)-p(u)|^4\right]\mathbb{E}^{\frac{1}{2}}\left[\frac{1}{|\widetilde{q}_n(u)|^4}\right] \le C\frac{1}{|q(u)|}.$$

Thanks to Lemma 6.4 and Lemma 6.2, for some C depending on Δ_{Max} , $C_4^{\pm}$ and κ,

$$\mathbb{E}^{\frac{1}{2}}\left[|\widehat{p}(u)-p(u)|^4\right]\mathbb{E}^{\frac{1}{2}}\left[\frac{1}{|\widetilde{q}(u)|^4}\right] \le C\frac{1}{|q(u)|}.$$

Consider now the second summand on the right hand side of formula (6.2). An application of Lemma 6.2, along with the fact that

$$|\Psi'(u)| = \left|\int xe^{iux}\nu(\mathrm{d}x)\right| \le \int |x|\nu(\mathrm{d}x) = \mathbb{E}\left[[]\,(X_+)_1\right]+\mathbb{E}\left[[]\,(X_-)_1\right] = C_1^{\pm},$$

yields for some C depending on Δ_{Max} and κ,

$$|p(u)|^2\mathbb{E}\left[\left|\frac{1}{q(u)}-\frac{1}{\widetilde{q}(u)}\right|^2\right] \le C\frac{|p(u)|^2}{|q(u)|^3} = C\frac{|\Psi'(u)|^2}{|q(u)|} \le C(C_1^{\pm})^2\frac{1}{|q(u)|}.$$

This completes the proof. □

6.2.2 Rate results: Upper bounds

Proof of Lemma 3.3. We can estimate

$$\|g - \mathrm{K}_h * g\|^2_{\mathrm{L}^2(\mathbb{R})} \leq 2\,\|\widetilde{g} - \mathrm{K}_h * \widetilde{g}\|^2_{\mathrm{L}^2(\mathbb{R})} + 2\,\|\mathrm{K}_h * (\widetilde{g} - g)\|^2_{\mathrm{L}^2(\mathbb{R})}\,.$$

A standard Taylor series argument implies for some constant C depending on the choice of K, on L, R and $\omega_2 - \omega_1$,

$$\|\widetilde{g} - \mathrm{K}_h * \widetilde{g}\|^2_{\mathrm{L}^2(\mathbb{R})} \leq Ch^{2a}.$$

It remains to consider the second summand. Let $\delta := \min\{\omega_1 - d_1, d_2 - \omega_2\}$. Using the fact that $\widetilde{g} - g$ vanishes on D, as well as (3.3), we have for arbitrary $x \in \Omega$:

$$\begin{aligned} &|\mathrm{K}_h * (\widetilde{g} - g)\,(x)| = \left|\int \frac{1}{h}\,\mathrm{K}\left(\frac{x-y}{h}\right)(\widetilde{g} - g)\,(y)\,\mathrm{d}y\right| \\ \leq\quad &\sup_{|z|\geq\delta} \frac{1}{h}\left|\mathrm{K}\left(\frac{z}{h}\right)\right|\left(\|\widetilde{g}\|_{\mathrm{L}^1(\mathbb{R})} + \|g\|_{\mathrm{L}^1(\mathbb{R})}\right) \leq \delta^{-2a-1}\left(\|\widetilde{g}\|_{\mathrm{L}^1(\mathbb{R})} + \|g\|_{\mathrm{L}^1(\mathbb{R})}\right)h^{2a}. \end{aligned}$$

This completes the proof. □

6.2.3 Minimax lower bounds

Proof of theorem 3.5 The minimax lower bounds are established by looking at a decision problem between an increasing finite number M of alternatives, see Theorem 2.5 in Tsybakov (2003). The construction of the alternatives essentially follows the proof of Theorem 4.4 in Neumann and Reiß (2009). For this reason, we only sketch the most important steps and omit some of the technical details.
Let

$$\eta(x) = b|x|e^{-\lambda|x|},\ x \in \mathbb{R}$$

be the Lévy density of a symmetric bilateral Gamma-distribution with parameters $b = \beta/2$ and with $\lambda > 0$ to be appropriately chosen. Let $g(x) := x\eta(x)$. By $\mathbb{P}_{0,j}$ we denote the infinitely divisible distribution with characteristic function

$$\varphi_{0,j}(u) = \exp\Big(\Delta_j \int (e^{iux} - 1)\eta(x)\,\mathrm{d}x\Big).$$

Since η is infinitely differentiable away from zero and $\eta(x) \sim |x|^{-1}$, $\eta \to 0$, one can, with the appropriate choice of λ, always guarantee that g or $g|_D$ is contained in the prescribed class of globally or locally regular functions.
By $f_{\mathcal{N}(\mu,\sigma)}$ we denote the density of a $\mathcal{N}_{\mu,\sigma}$-distribution. For $n \in \mathbb{N}$, a positive integer m_n to be appropriately chosen and some universal positive constant $d > 0$, we introduce the following perturbations of η,

$$h_{n,j}(x) := dK_n 2\sin((m_n + j)x)f_{\mathcal{N}(0,1)}(x),\ j = 0, \cdots, m_n - 1.$$

When locally Hölder regular functions are considered, $K_n := m_n^{-a-\frac{1}{2}}$. For functions g having a Fourier transform which decays at the rate $|u|^{-1}$, $K_n := m_n^{-1}$. For $S \subseteq \{0, \cdots, m_n - 1\}$, we define

$$\eta_{S,n}(x) := \eta(x) + \sum_{j\in S} h_{n,j}(x).$$

Moreover, $g_{S,n}(x) := x\eta_{S,n}(x)$. Using the fact that the multiplication with a sine- function corresponds to a shift in the Fourier domain, we find that

$$\begin{aligned}&\mathcal{F}(xh_{n,j}(x))(u)\\ =&\frac{dK_n}{2}\Big[(u-(m_n+j))\exp\Big(-\frac{1}{2}(u-(m_n+j))^2\Big)-(u+(m_n+j))\exp\Big(-\frac{1}{2}(u+m_n+j)^2\Big].\end{aligned}$$

On the other hand, arguments presented in Neumann and Reiß (2009) ensure that the characteristic function of the perturbed distribution has the same decay behavior as $\varphi_{0,j}$.
From there, we derive that the perturbed functions $g_{S,n}$ are still in the prescribed classes of globally or locally regular functions.
Let us estimate, for $S_1, S_2 \subseteq \{0, \cdots, m_n-1\}$, the L^2-distance between $g_{S_1,n}$ and $g_{S_2,n}$. We start by observing that, using elementary calculus, the decay of the normal density as well as the periodicity of the sine-function,

$$\begin{aligned}\int_\Omega |g_{S_1,n}(x)-g_{S_2,n}(x)|^2\,\mathrm{d}x &= \int_\Omega \Big|\sum_{j\in S_1\triangle S_2} xh_{h,j}(x)\Big|^2\,\mathrm{d}x\\ &\geq c\int\Big|\sum_{j\in S_1\triangle S_2} xh_{h,j}(x)\Big|^2\,\mathrm{d}x = c\int|g_{S_1,n}(x)-g_{S_2,n}(x)|^2\,\mathrm{d}x\end{aligned}$$

for a small enough positive constant c. By the Plancherel formula and the construction of the $h_{n,j}$,

$$\begin{aligned}&\int|g_{S_1,n}(x)-g_{S,n}(x)|^2\,\mathrm{d}x = \frac{1}{2\pi}\int|\sum_{j\in S_1\triangle S_2}\mathcal{F}(xh_{n,j}(x))(u)|^2\,\mathrm{d}u\\ \geq&\frac{1}{2\pi}\sum_{j\in S_1\triangle S_2}\int|\mathcal{F}(xh_{n,j}(x))(u)|^2\,\mathrm{d}u \geq C|S_1\triangle S_2|K_n^2.\end{aligned}$$

For some universal positive constant C.
Next, we calculate the Kullback-Leibler divergence between the competing measures.

$$\mathrm{KL}(\mathbb{P}_{S,n}\,|\,\mathbb{P}_{0,n}) = \sum_{j=1}^n \mathrm{KL}(\mathbb{P}_{S,n,j}\,|\,\mathbb{P}_{0,n,j}) \leq \sum_{j=1}^n \chi(\mathbb{P}_{S,n,j},\mathbb{P}_{0,n,j}).$$

We may again argue along the same lines as in the proof of Theorem 4.4 in Neumann and Reiß (2009) to find that for some positive constant C' depending on d,

$$\sum_{j=1}^n \chi(\mathbb{P}_{S,n,j},\mathbb{P}_{0,n,j}) \leq C'|S|\sum_{j=1}^n \Delta_j(1+m_n)^{-2\Delta_j\beta}K_n^2.$$

For $m_n \geq 8$, the Varshamov-Gilbert Lemma (see chapter 2 in Tsybakov (2003)) guarantees the existence of $M \geq \sqrt[8]{2}^{m_n}$ different subsets $S_0, \cdots, S_{M-1}$ of $\{0, \cdots, m_n-1\}$, including the empty set $S_0 = \emptyset$, such that

$$|S_k \triangle S_\ell| \geq m_n/8 \quad \forall 0 \leq k < \ell \leq M-1.$$

Theorem 2.5 in Tsybakov (2003) then implies that, with K_n implicitly defined by $K_n \sum_{j=1}^{n} \Delta_j(1+|m_n|)^{-2\beta} = 1$,

$$\limsup_{n\to\infty} \inf_{\widetilde{g}_n} \sup_{\breve{g}\in\mathcal{G}} \mathbb{E}_g\left[\|\breve{g}-g\|_{\mathrm{L}^2}^2\right] K_n^{-2} m_n \geq 0.$$

Here $\mathcal{G}$ may either stand for the class $\mathcal{G}_{\mathrm{pol}}(\beta, C_\varphi, c_\varphi, C_g, C)$ or $\mathcal{G}(a, D, L, R, C_1, \beta, C_2, C_3)$ and L^2 denotes the L^2 norm on the whole real axes or on Ω, respectively. This completes the proof. □

6.3 Estimating the distributional density

6.3.1 Upper risk bound

Before proving the main result of the section, we formulate an auxiliary result. The proof of the Lemma is postponed to the appendix.

Lemma 6.5. *For any $m \in \mathbb{N}$, there exists a constant C depending on m, Δ_{max} and $C_{4m}^{\pm}$ such that*

$$\mathbb{E}\Big[\Big|\int_0^u \frac{\widehat{p}(x)}{\widetilde{q}(x)} - \frac{p(x)}{q(x)}\,\mathrm{d}x\Big|^{2m}\Big] \leq C\Big(C_{\Psi,1}\int_0^u \frac{1}{q(x)}\,\mathrm{d}x\Big)^m \Big(\max\Big\{1, C_{\Psi,2}^2\int_0^u \frac{1}{q(x)}\,\mathrm{d}x\Big\}\Big)^m,$$

with

$$C_{\Psi,1} = (\|\Psi'\|_{\mathrm{L}^2,u}^2 + \|\Psi''\|_{\mathrm{L}^1,u}) \vee 1 \quad \textit{and} \quad C_{\Psi,2} = \|\Psi'\|_{\infty,u} \vee 1.$$

Proof of Theorem 4.1 Parseval's identity, along with the fact that K is compactly supported gives

$$\mathbb{E}\big[\|f - \widehat{f}_{h,n}\|_{\mathrm{L}^2}^2\big] \leq 2\|f - \mathrm{K}_h * f\|_{\mathrm{L}^2}^2 + \frac{1}{\pi}\int_{-1/h}^{1/h} \mathbb{E}\big[|\widehat{\varphi}(u) - \varphi(u)|^2\big]\,\mathrm{d}u$$

In the sequel, we use the short notation notation

$$\Delta(u) := \int_0^u \left(\frac{\widehat{p}(z)}{\widetilde{q}(z)} - \frac{p(z)}{q(z)}\right)\mathrm{d}z.$$

Since we have, by definition of $\widehat{\varphi}$, $|\widehat{\varphi}(u) - \varphi(u)| \leq |\breve{\varphi}(u) - \varphi(u)|$, as well as $|\widehat{\varphi}(u) - \varphi(u)| \leq 2$, we can estimate

$$|\widehat{\varphi}(u) - \varphi(u)|^2 \leq |\breve{\varphi}(u) - \varphi(u)|^2 \mathbb{1}_{\{|\Delta(u)|\leq 1\}} + 4\mathbb{1}_{\{|\Delta(u)|>1\}}$$

Using the definition of $\breve{\varphi}$, as well as the fact that $|1 - \exp(z)| \leq 2|z|$ holds for $|z| \leq 1$, we can continue by estimating for arbitrary $m \in \mathbb{N}$:

$$\begin{aligned}|\breve{\varphi}(u) - \varphi(u)|^2\mathbb{1}_{\{|\Delta(u)|\leq 1\}} + 4\mathbb{1}_{\{|\Delta(u)|>1\}} &\leq |\varphi(u)|^2|1 - \exp(\Delta(u))|^2\mathbb{1}_{\{|\Delta(u)|\leq 1\}} + 4\mathbb{1}_{\{|\Delta(u)|>1\}}\\ &\leq 4|\varphi(u)|^2|\Delta(u)|^2 + 4|\Delta(u)|^m.\end{aligned}$$

By Lemma 6.5, for some C depending on Δ_{Max} and on $\mathbb{E}\left[X_{4m}^{\pm}\right]$,

$$\int_{-1/h}^{1/h} |\varphi(u)|^2 \mathbb{E}\big[|\Delta(u)|^2 \,\mathrm{d}u\big] \le C C_{\Psi,1} \int_{-1/h}^{1/h} |\varphi(u)|^2 \int_0^u \frac{1}{q(z)} \,\mathrm{d}z \,\mathrm{d}u$$

and

$$\int_{-1/h}^{1/h} \mathbb{E}\big[|\Delta(u)|^m\big] \,\mathrm{d}u \le C \int_{-1/h}^{1/h} \Big(C_{\Psi,1} \int_0^u \frac{1}{q(x)} \,\mathrm{d}x \Big)^m \Big(\max\Big\{1, C_{\Psi,2} \int_0^u \frac{1}{q(z)} \,\mathrm{d}z \Big\}\Big)^m \,\mathrm{d}u.$$

This completes the proof. □

Appendix

Proof of Lemma 6.5 We use the estimate

$$\begin{aligned}
&\Big| \int_0^u \frac{\widehat{p}(x)}{\widetilde{q}(x)} - \frac{p(x)}{q(x)} \,\mathrm{d}x \Big|^{2m} \\
\le& 3^{2m-1} \Big(\Big| \int_0^u \frac{\widehat{p}(x) - p(x)}{q(x)} \,\mathrm{d}x \Big|^{2m} + \Big| \int_0^u (\widehat{p}(x) - p(x)) \mathrm{R}(x) \,\mathrm{d}x \Big|^{2m} + \Big| \int_0^u p(x) \mathrm{R}(x) \,\mathrm{d}x \Big|^{2m} \Big)
\end{aligned} \tag{6.13}$$

We bound, successively, the expected value of each of the three terms in the second line of formula ().

- Consider the first summand: By the Rosenthal inequality,

$$\begin{aligned}
&\mathbb{E}\Big[\Big| \int_0^u \frac{\widehat{p}(x) - p(x)}{q(x)} \,\mathrm{d}x \Big|^{2m}\Big] \\
\le& \max\Big\{ \sum_{j=1}^n \mathbb{E}\Big[\Big| \int_0^u \frac{w_j(x)(\varphi_j' - \widehat{\varphi}_j')(x)}{q(x)} \,\mathrm{d}x \Big|^{2m}\Big], \Big(\sum_{j=1}^n \mathbb{E}\Big[\Big| \int_0^u \frac{w_j(x)(\varphi_j' - \widehat{\varphi}_j')(x)}{q(x)} \,\mathrm{d}x \Big|^2\Big]\Big)^m \Big\}
\end{aligned} \tag{6.14}$$

For any integer $k \ge 2$,

$$\begin{aligned}
&2^{-k} \mathbb{E}\Big[\Big| \int_0^u \frac{w_j(x)(\widehat{\varphi}_j' - \varphi_j')(x)}{q(x)} \,\mathrm{d}x \Big|^k\Big] \le \mathbb{E}\Big[\Big| \int_0^u \frac{w_j(x)\widehat{\varphi}_j'(x)}{q(x)} \,\mathrm{d}x \Big|^k\Big] \\
=& \int_{[0,u]^k} \prod_{\ell=1}^k \frac{w_j\big((-1)^{\ell+1} x_\ell\big)}{q\big((-1)^{\ell+1} x_\ell\big)} \mathbb{E}\Big[(iZ_j)^k \exp\Big(i\Big(\sum_{\ell=1}^k (-1)^{\ell+1} x_\ell\Big) Z_j\Big)\Big] \,\mathrm{d}x_1 \cdots \mathrm{d}x_k \\
=& \int_{[0,u]^k} \prod_{\ell=1}^k \frac{w_j\big((-1)^{\ell+1} x_\ell\big)}{q\big((-1)^{\ell+1} x_\ell\big)} \varphi_j^{(k)}\Big(\sum_{\ell=1}^k (-1)^{\ell+1} x_\ell\Big) \,\mathrm{d}x_1 \cdots \mathrm{d}x_k.
\end{aligned}$$

Repeated applications of the integration by parts formula give for some constant C depending on k,

$$
\begin{aligned}
&\int\limits_{[0,u]^k} \prod_{\ell=1}^{k} \frac{w_j\big((\text{-}1)^{\ell+1}x_\ell\big)}{q\big((\text{-}1)^{\ell+1}x_\ell\big)} \varphi_j^{(k)}\Big(\sum_{\ell=1}^{k}(\text{-}1)^{\ell+1}x_\ell\Big)\,\mathrm{d}x_1\cdots\mathrm{d}x_k \\
\leq & C\sum_{m\leq k-2}\Big(\max\Big\{\frac{1}{q(u)},\frac{1}{q(0)}\Big\}^m \sup_{t\in\mathbb{R}}\int_0^u\int_0^u \frac{|\varphi_j''(t+x_1)w_j(x_1)w_j(x_2)|}{|q(x_1)q(x_2)|}\,\mathrm{d}x_1\,\mathrm{d}x_2 \\
&\qquad \int\limits_{[0,u]^{k-m-2}} \prod_{\ell=2}^{k-m-2}\Big|\Big(\frac{w_j}{q}\Big)'(x_\ell)\Big|\,\mathrm{d}x_3\cdots\mathrm{d}x_{k-m-1}\Big).
\end{aligned}
$$

First, using $\varphi_j'' = \Delta_j\Psi''\varphi_j + \Delta_j^2(\Psi')^2\varphi_j$ and the Cauchy-Schwarz inequality, we obtain

$$
\int_0^u\int_0^u \frac{|\varphi_j''(t+x_1)w_j(x_1)w_j(x_2)|}{|q(x_1)q(x_2)|}\,\mathrm{d}x_1\,\mathrm{d}x_2 \leq (\|\Psi'\|_{\mathrm{L}^2}^2 + \|\Psi''\|_{\mathrm{L}^1})\Delta_{\mathrm{Max}}\int_0^u \frac{\Delta_j|w_j(x)|^2}{q(x)^2}\,\mathrm{d}x.
$$

On the other hand,

$$
q'(x) = \sum_{j=1}^{n}\Delta_j(w_j'(x)\varphi_j(x) + w_j(x)\varphi_j'(x)) = 2\sum_{j=1}^{n}\Delta_j^2\Psi'(x)|\varphi_j(x)|^2
$$

and consequently,

$$
\Big|\Big(\frac{w_j}{q}\Big)'(x)\Big| \leq \frac{w_j'(x)}{q(x)} + 2\Delta_{max}\frac{\Psi'(x)w_j(x)}{q(x)} = (1+2\Delta_{max})\frac{\Psi'(x)w_j(x)}{q(x)}.
$$

This implies

$$
\int\limits_{[0,u]^{k-m-2}} \prod_{\ell=3}^{k-m-2}\Big|\Big(\frac{w_j}{q}\Big)'(x_\ell)\Big|\,\mathrm{d}x_3\cdots\mathrm{d}x_{k-m-2} \leq \Big(3\Delta_{\mathrm{Max}}\|\Psi'\|_\infty\int_0^u \frac{1}{q(x)}\,\mathrm{d}x\Big)^{k-m-2}.
$$

Finally, using the fact that q is bounded above by T and $q(0) = T$, we find that for any $u \geq 1$,

$$
\max\Big\{\frac{1}{q(u)},\frac{1}{q(0)}\Big\} \leq \frac{2}{q(0)} + \Delta_{max}\int_0^u \frac{|\Psi'(x)|}{q(x)}\,\mathrm{d}x \leq \Delta_{\mathrm{Max}}(|\Psi'\|_\infty + 2)\int_0^u \frac{1}{q(x)}\,\mathrm{d}x.
$$

Putting the above together, we have shown that for any integer $k \geq 2$, there exists a constant C depending on Δ_{max} and k such that

$$
\mathbb{E}\Big[\Big|\int\limits_0^u \frac{w_j(x)\widehat{\varphi}_j'(x)}{q(x)}\,\mathrm{d}x\Big|^k\Big] \leq C\Big(C_{\Psi,2}\int_0^u \frac{1}{q(x)}\,\mathrm{d}x\Big)^{k-2} C_{\Psi,1}\int_0^u \frac{\Delta_j|w_j(x)|^2}{q(x)^2}\,\mathrm{d}x.
$$

From there, we can conclude that

$$\sum_{j=1}^{n}\mathbb{E}\Big[\Big|\int_0^u \frac{w_j(x)(\widehat{\varphi}_j'-\varphi_j')(x)}{q(x)}\,\mathrm{d}x\Big|^k\Big]\leq C\Big(C_{\Psi,2}\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big)^{k-2}C_{\Psi,1}\int_0^u\frac{\sum_{j=1}^n\Delta_j|w_j(x)|^2}{q(x)^2}\,\mathrm{d}x$$

$$=CC_{\Psi,1}C_{\Psi,2}^{k-2}\Big(\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big)^{k-1}.$$

Combining this with (), we have shown that for any $m\in\mathbb{N}$, there exists a constant C depending on m and Δ_{max} such that

$$\mathbb{E}\Big[\Big|\int_0^u\frac{\widehat{p}(x)-p(x)}{q(x)}\,\mathrm{d}x\Big|^{2m}\Big]\leq C\Big(C_{\Psi,1}\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big)^m\Big(\max\Big\{1,C_{\Psi,2}^2\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big\}\Big)^{m-1}.$$

-Consider now the second summand in (). We apply the Hölder inequality, Lemma 6.2 and Lemma 6.4 to derive that for a constant C depending on m, Δ_{Max} and $C_{4m}^{\pm}$,

$$\mathbb{E}\Big[\Big|\int_0^u(\widehat{p}(x)-p(x))\mathrm{R}(x)\,\mathrm{d}x\Big|^{2m}\Big]\leq\Big(\int_0^u\mathbb{E}\big[|\widehat{p}(x)-p(x)|^{4m}\big]^{\frac{1}{4m}}\mathbb{E}\big[|\mathrm{R}(x)|^{4m}\big]^{\frac{1}{4m}}\,\mathrm{d}x\Big)^{2m}$$
$$\leq C\Big(\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big)^{2m}.$$

-Finally, the third summand can be bounded as follows: By the definition of p, the Cauchy-Schwarz inequality and Lemma (6.2),

$$\mathbb{E}\Big[\Big|\int_0^u p(x)\mathrm{R}(x)\,\mathrm{d}x\Big|^{2m}\Big]\leq\Big(\int_0^u|p(x)|\mathbb{E}\Big[|\mathrm{R}(x)|^{2m}\Big]^{\frac{1}{2m}}\,\mathrm{d}x\Big)^{2m}\leq C\Big(\int_0^u\frac{|\Psi'(x)q(x)|}{q(x)^{\frac{3}{2}}}\,\mathrm{d}x\Big)^{2m}$$

$$\leq C\|\Psi'\|_{\mathrm{L}^2}^{2m}\Big(\int_0^u\frac{1}{q(x)}\,\mathrm{d}x\Big)^m.$$

This completes the proof. □

Acknowledgement.

The author is grateful to Alexander Meister for intense discussions on this paper and for useful hints and ideas both, on the theoretical part and on the numerical procedure.

References

Ishwar V. Basawa and Peter J. Brockwell. Non-parametric estimation for non-decreasing Lévy Processes. *Journal of the Royal Statistical Society. Series B*, 44(2):262–269, 1982.

Denis Belomestny. Statistical infernece for time-changed Lévy processes via composite characteristic function estimation. *The Annals of Statistics*, 39(4):2205–2242, 2011.

Denis Belomestny and Mathias Trabs. Low-rank diffusion matrix estimation for high-dimensional time-changed Lévy processes. arXiv:1510.04638v1, 2015.

Fabienne Comte and Valentine Genon-Catalot. Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, 119: 4088–4123, 2009.

Fabienne Comte and Valentine Genon-Catalot. Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'Institut Henri Poincaré*, 3(46):595–617, 2010.

Fabienne Comte and Valentine Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *Annals of Statistics*, 39(2):803–837, 2011.

Aurore Delaigle and Alexander Meister. Density estimation with heteroscedastic error. *Bernoulli*, 14(2):562–579, 2008.

José E. Figueroa-López. Nonparametric estimation for Lévy models based on discrete sampling. *IMS Lecture Notes of the 3rd E.L. Lehmann Symposium*, 57:117–146, 2009.

Rustam Ibragimov and Shaturgun Sharakhmetov. The exact constant in the Rosenthal inequality for random variables with mean zero. *Theory of Probability and Its Applications*, 46(1): 127–132, 2002.

Johanna Kappus. Adaptive nonparametric estimation for lévy processes observed at low frequency. *Stochastic Processes and their Applications*, 124(1):730–758, 2014.

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

Michael H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.

Steven Orey. On continuity properties of infinitely divisible functions. *The Annals of Mathematical Statistics*, 39(3):936–937, 1968.

Herman Rubin and Howard G. Tucker. Estimating the parameters of a differential process. *The Annals of Mathematical Statistics*, 30(3):641–658, 1959.

Ken-Iti Sato. *Lévy processes and infinitely divisible distributions.* Cambridge University Press, 1999.