Ester Mariucci

Distances for discretely observed jump processes and applications in nonparametric statistics

Joint work with Alexandra Carpentier, Céline Duval, Markus Reiß
SFB colloquium, 22 November 2019

Institut für Mathematische Optimierung
Otto-von-Guericke-Universität Magdeburg

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

DFG-Graduiertenkolleg
MATHEMATISCHE KOMPLEXITÄTSREDUKTION

Table of contents

1. Equivalence in Le Cam sense
2. Inference for the Lévy density: the general problem
3. Inference via Le Cam theory
4. Inference via a direct approach
5. Gaussian approximation of the small jumps
6. Conclusion

Equivalence in Le Cam sense

Problem: When do two statistical models

$$\mathcal{E}_1 = \big(\mathcal{X}_1, \mathcal{T}_1, (P_{1,\theta} : \theta \in \Theta)\big) \quad \text{and} \quad \mathcal{E}_2 = \big(\mathcal{X}_2, \mathcal{T}_2, (P_{2,\theta} : \theta \in \Theta)\big)$$

contain the same amount of information about θ?

Equivalence in Le Cam sense

Problem: When do two statistical models

$$\mathcal{E}_1 = \big(\mathcal{X}_1, \mathcal{T}_1, (P_{1,\theta} : \theta \in \Theta)\big) \quad \text{and} \quad \mathcal{E}_2 = \big(\mathcal{X}_2, \mathcal{T}_2, (P_{2,\theta} : \theta \in \Theta)\big)$$

contain the same amount of information about θ?

Idea: The statistical models $\mathcal{E}_1$ and $\mathcal{E}_2$ contain "the same amount of information about any $\theta \in \Theta$" if there exist two Markov kernels, K_1 and K_2 that do not depend on θ, such that

$$K_1 P_{1,\theta} = P_{2,\theta} \quad \text{and} \quad K_2 P_{2,\theta} = P_{1,\theta} \quad \forall \theta \in \Theta.$$

Equivalence in Le Cam sense

Problem: When do two statistical models

$$\mathcal{E}_1 = \big(\mathcal{X}_1, \mathcal{T}_1, (P_{1,\theta} : \theta \in \Theta)\big) \quad \text{and} \quad \mathcal{E}_2 = \big(\mathcal{X}_2, \mathcal{T}_2, (P_{2,\theta} : \theta \in \Theta)\big)$$

contain the same amount of information about θ?

Idea: The statistical models $\mathcal{E}_1$ and $\mathcal{E}_2$ contain "the same amount of information about any $\theta \in \Theta$" if there exist two Markov kernels, K_1 and K_2 that do not depend on θ, such that

$$K_1 P_{1,\theta} = P_{2,\theta} \quad \text{and} \quad K_2 P_{2,\theta} = P_{1,\theta} \quad \forall \theta \in \Theta.$$

Definition

The **Le Cam distance** $\Delta\big(\mathcal{E}_1, \mathcal{E}_2\big)$ is defined as

$$\delta(\mathcal{E}_1, \mathcal{E}_2) = \inf_K \sup_{\theta \in \Theta} \| K P_{1,\theta} - P_{2,\theta} \|_{TV}, \quad \Delta(\mathcal{E}_1, \mathcal{E}_2) = \max(\delta(\mathcal{E}_1, \mathcal{E}_2), \delta(\mathcal{E}_2, \mathcal{E}_1)),$$

where the infimum is taken over all Markov kernels.

Le Cam theory and decision theory

How to transfer estimators: Suppose that $KP_{1,\theta} = P_{2,\theta}$, with $K(x,A) = \mathbf{1}_A S(x)$ for all $x \in \mathcal{X}_1$, $A \in \mathcal{T}_2$ and $\theta \in \Theta$. For any estimator π_2 in $\mathscr{E}_2$ it is possible to define an estimator π_1 in $\mathscr{E}_1$ with the same risk as π_2. Simply take

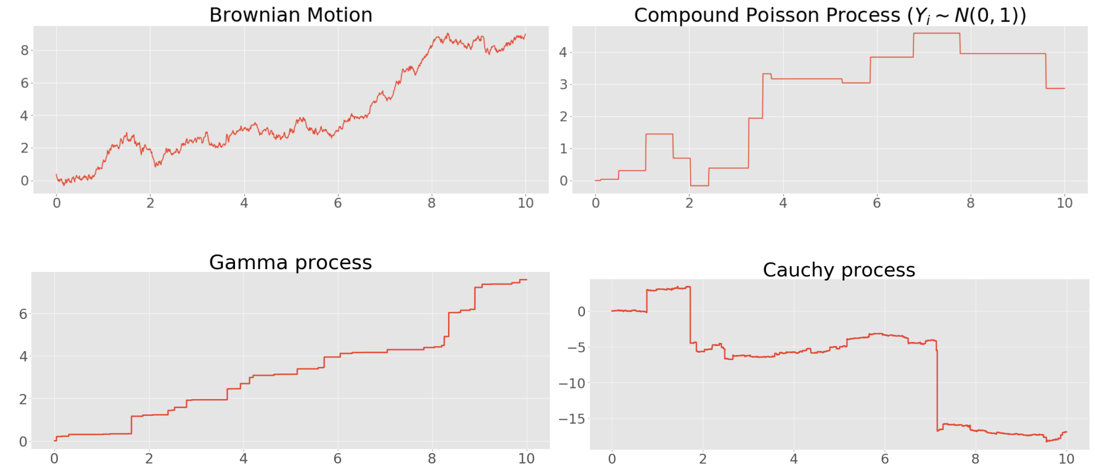
$$\pi_1(x) := \pi_2(S(x)), \forall x \in \mathcal{X}_1.$$

Philosophy: If two models are equivalent then they have the same statistical properties $\Longrightarrow$ it is enough to choose the simplest one when studying these properties.

L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York. (1986).

Lévy processes

Lévy processes are stochastic processes with independent, stationary increments and trajectories a.s. càdlàg (right continuous with left limits).



Why Lévy processes?

- Lévy processes are the fundamental building blocks in stochastic models with evolution in time exhibiting sudden changes in value.
- They play a central role in many fields of science such as:
  - *mathematical finance*: for modelling market fluctuations;
  - *actuarial science*: for risk theory and premium calculations;
  - *biology*: for modelling the membrane potential;
  - *hydrology*: for modelling rainfall;
  - *physics*: for weather and climate;
  - *seismology*: for earthquakes;
  - *engineering*: to model the functioning of GPS;
  - many others...

Structure of a Lévy process

Lévy process = Gaussian process + small jumps + compound Poisson process.

Structure of a Lévy process

Lévy process = Gaussian process + small jumps + compound Poisson process.

Formally, $X \sim (b, \Sigma^2, \nu)$ can be written as:

$$X_t = \Sigma W_t + bt + X_t^S + X_t^B,$$

- W is a standard Brownian motion;
- X^S is a centred martingale describing the small jumps;
- X^B is a compound Poisson process: $X_t^B = \sum_{i=1}^{N_t} Y_i$ with N a Poisson process of intensity $\lambda := \nu(\mathbb{R} \setminus [-1, 1])$ independent of the sequence of i.i.d. $(Y_i)_{i \geq 0}$, $Y_1 \sim \nu_{|\mathbb{R} \setminus [-1,1]} / \lambda$;
- W, X^S and X^B are independent of each other.

Interpretation of the Lévy measure

The jump dynamics of a Lévy process is determined by its Lévy measure ν.
Interpretation of ν: $\forall B \in \mathscr{B}(\mathbb{R} \setminus \{0\})$

$$\nu(B) = \frac{1}{t}\mathbb{E}\Bigg[\sum_{0<s\leq t} \mathbf{1}_B(\Delta X_s)\Bigg] \tag{1}$$

is the average number of jumps whose size is in B.

Example Let $B = (0,1]$ and X be a Lévy process with Lévy density $\frac{\nu(dx)}{dx} = \frac{e^{-x}}{x}\mathbf{1}_{x>0}$. In particular, $\nu(B) = \infty$. Thus, from (1), we deduce that for any time interval $[0,t]$, $(X_s)_{s\in[0,t]}$ will have in expectation infinitely many jumps with size in $(0,1]$.

Lévy density estimation from high frequency data

Let X be a pure jump Lévy process with Lévy measure ν.

Data: $(X_{i\Delta})_{i=1}^n$, with $\Delta = \Delta_n \to 0$ and $n\Delta_n \to \infty$ as $n \to \infty$.

Assumption: ν is absolutely continuous with respect to a dominating measure ν_0. The density $f = \frac{d\nu}{d\nu_0}$ belongs to a nonparametric class $\mathscr{F}$.

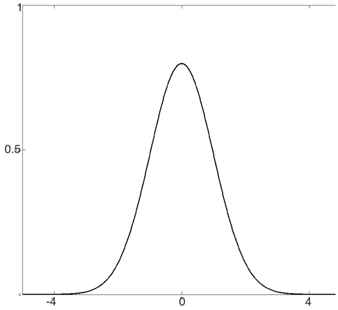
Goal: Build minimax estimators for the Lévy density $f = d\nu/d\nu_0$, i.e. estimators that minimise the minimax risk with respect to a loss function ℓ:

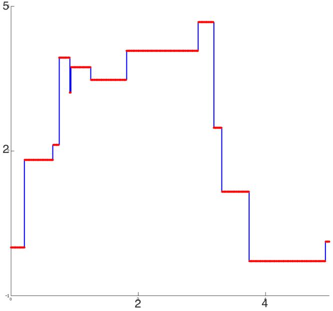
$$R^* := \inf_T \sup_{f \in \mathscr{F}} \mathbb{E}[\ell(T, f)].$$

Remark

1. We face an inverse problem: we do not observe samples of ν but of the law of the increments of X.
2. There is no explicit, clear link between the law of X and ν.
3. ν is an infinite dimensional object. We face a nonparametric estimation problem.
4. If $\nu(\mathbb{R}) = \infty$ and $\nu_0 = Leb$, the Lévy density explodes in a neighborhood of zero.

Illustration: Compound Poisson process $\lambda = 2$, $\Delta = 0.01$

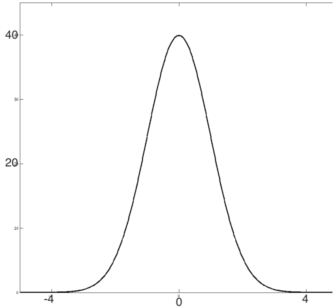




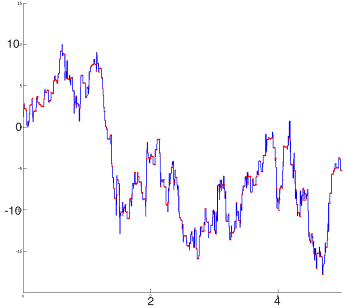
Lévy density: $f(x) := \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,
$\nu(\mathbb{R}) = 2, \ \nu_0 = Leb$

One trajectory of X

Illustration: Compound Poisson process $\lambda = 100$, $\Delta = 0.01$

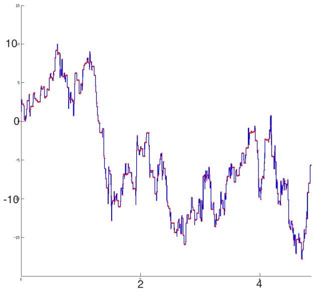


Lévy density: $f(x) := \frac{100}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,
$\nu(\mathbb{R}) = 100$, $\nu_0 = Leb$

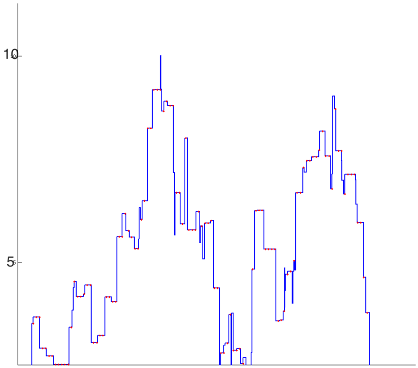


One trajectory of X

Illustration: Compound Poisson process $\lambda = 100$, $\Delta = 0.01$



One trajectory of X



Zoom

Complexity reduction via approximations

$$X = \text{ Gaussian process} + \text{ small jumps } + \text{ compound Poisson process.}$$

- **Le Cam theory:** consists of finding a statistical model, simpler than the Lévy one, and asymptotically equivalent to it for what concerns the estimation of f.
- **Compound Poisson approximation:** consists of ignoring the small jumps. The theoretical justification comes from the fact that any Lévy process is the (weak) limit of a sequence of compound Poisson processes.
- **Gaussian approximation for the small jumps:** consists of proving that the law of the discrete observations of the small jumps strongly converges to the law of a Gaussian vector.

Literature

Case $\nu(\mathbb{R}) < \infty$: well understood.

Case $\nu(\mathbb{R}) = \infty$: more challenging because of the small jumps. The analysis so far relies on spectral approaches (L_2 and L_∞ loss) and one typically estimates functionals of f such as $xf(x)$ and $x^2 f(x)$.

Literature

Case $\nu(\mathbb{R}) < \infty$: well understood.

Case $\nu(\mathbb{R}) = \infty$: more challenging because of the small jumps. The analysis so far relies on spectral approaches (L_2 and L_∞ loss) and one typically estimates functionals of f such as $xf(x)$ and $x^2 f(x)$.

References (from high- and low-frequency observations):
Compound Poisson process: Buchmann and Grübel, (2003); van Es, Gugushvili, Spreij, (2007); Duval, (2013); Comte, Duval, Genon-Catalot, (2014); Coca, (2018); Gugushvili, M., van der Meulen (2019).

Infinite activity: Comte, Genon-Catalot, (2009), (2010), (2011); Figueroa-López, Houdré, (2009); Neumann, Reiß, (2009); Kappus, Reiß, (2010); Gugushvili, (2012); Trabs, (2015).

Textbook: Lévy matters IV - Estimation for discretely observed Lévy processes. Belomenstny, Comte, Genon-Catalot, Masuda, Reiß, (2015).

Asymptotic equivalence result

$$\mathscr{P}_n: \quad (X_{i\Delta} - X_{(i-1)\Delta})_{i=1}^n, \quad X = X^S + X^B \text{ with } f = d\nu/d\nu_0 \text{ on } I, \quad f \in \mathscr{F},$$

$$\mathscr{W}_n: \quad dy_t = \sqrt{f(t)}dt + \frac{dW_t}{2\sqrt{n\Delta}\sqrt{g(t)}}, \quad g = \frac{d\nu_0}{dx}, \quad t \in I, \quad f \in \mathscr{F}.$$

Theorem (M., 2016)

Suppose that $\Delta = \Delta_n \to 0$, $n\Delta_n \to \infty$ *and* $n\Delta_n^2 \to 0$ *as* $n \to \infty$. *Under appropriate assumptions on* $\mathscr{F}$ *it holds:*

$$\Delta(\mathscr{P}_n, \mathscr{W}_n) \to 0 \quad \text{as} \quad n \to \infty.$$

The upper bound on $\Delta(\mathscr{P}_n, \mathscr{W}_n)$ *is explicit as well as all the equivalence mappings.*

E. Mariucci, Asymptotic equivalence for pure jump Lévy processes with unknown Lévy density and Gaussian white noise. Stoch. Proc. Appl. 126.2 (2016)

Interpretation

It is well known that a density estimation problem with i.i.d. random variables $(X_i)_{i=1}^n$ with support in I and common density $f \in \mathscr{F}$ (with respect to the Lebesgue measure) is asymptotically equivalent to a Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{dW_t}{2\sqrt{n}}, \quad t \in I, \quad f \in \mathscr{F}.$$

$\Longrightarrow$ Estimating the Lévy density from the observation of n equidistant increments of a Lévy process with a sample rate Δ is "as difficult" as estimating the density of $n\Delta$ i.i.d. random variables.

$\Longrightarrow$ If $\mathscr{F}$ is a class of regularity s, then we expect that the minimax rate of convergence for the estimation of the Lévy density $f \in \mathscr{F}$ is $(n\Delta)^{-\frac{sp}{2s+1}}$ for an L_p loss.

Compound Poisson approximation

- Let $I(\varepsilon) := (-\overline{A}, -\varepsilon] \cup [\varepsilon, \overline{A})$, $\varepsilon > 0$ and $\overline{A} \in (\varepsilon, \infty]$. For any $x \in I(\varepsilon)$,

$$f(x) = f(x)\mathbf{1}_{|x|>\varepsilon} = \lambda_\varepsilon h_\varepsilon(x),$$

  where $h_\varepsilon(x) = \frac{f(x)}{\lambda_\varepsilon}\mathbf{1}_{|x|>\varepsilon}$.
- We observe that $\lambda_\varepsilon h_\varepsilon(x)$ is the Lévy density of a CPP with intensity $\lambda_\varepsilon := \nu(x : |x| > \varepsilon)$ and jump density h_ε.
- We define

$$\hat{f}_n(x) := \hat{\lambda}_{n,\varepsilon}\hat{h}_{n,\varepsilon}(x), \quad \forall x \in I(\varepsilon).$$

- We assume $f \in \mathscr{F}(s, p, q, M_\varepsilon)$,

$$\mathscr{F}(s, p, q, M_\varepsilon,) = \left\{f \in L_p(I(\varepsilon)) \ : \ \|f\|_{B^s_{p,q}(I(\varepsilon))} \leq \mathcal{M}_\varepsilon\right\},$$

  where $\mathcal{M}_\varepsilon := \lambda_\varepsilon \mathcal{M}$, $\mathcal{M}$ being fixed, $p \in [2, \infty)$, $q \in [1, \infty]$ and $s > 1/p$.

Main result (in a simplified setting)

Set $\mathscr{F} = \mathscr{F}(s,p,q,M_\varepsilon) \cap \left\{ f : f(x) \leq \frac{M}{|x|^{1+\alpha}} \ \forall |x| \leq 2 \right\} \cap \left\{ f : \int_{|x|\geq\varepsilon} f(x)dx \geq 1 \right\}$.

Theorem (Duval, M.)

Let X be a pure jump Lévy process with Lévy density f. Fix $\varepsilon \in (0,1]$ and suppose that $f \in \mathscr{F}$ for some $p \in [2,\infty)$, $q \in [1,\infty]$, $s \geq 3/2 - 1/p$, $M_\varepsilon, M > 0$ and $\alpha \in (0,1)$. Then, for all $n \geq 1$ and $\Delta > 0$ such that $n\Delta \geq 1$ and $\Delta \leq \mathcal{C}_\alpha \varepsilon^\alpha M^{-1}$, it holds that

$$\sup_{f\in\mathscr{F}} \mathbb{E}\left[\int_{I(\varepsilon)} |\hat{f}_{n,\varepsilon}(x) - f(x)|^p dx \right] \leq C(n\Delta)^{-\frac{sp}{2s+1}},$$

where $C > 0$ is a constant independent of n and Δ.

Remarks

It is possible to have a more general upper bound for $\mathbb{E}\big[\int_{I(\varepsilon)} |\hat{f}_{n,\varepsilon}(x) - f(x)|^p dx\big]$ that holds true for all $\varepsilon \in (0,1]$, $f_\varepsilon \in \mathscr{F}$, $\Delta > 0$, $n\Delta \geq 1$. From such a bound one derives that:

- The estimator $\hat{f}_{n,\varepsilon}$ is consistent as soon as $n\Delta \to \infty$.
- Up to a log-factor, the rate of convergence for $\hat{f}_{n,\varepsilon}$ is still $(n\Delta)^{-\frac{sp}{2s+1}}$ even for infinite variation Lévy processes with a Lévy measure controlled by Mx^{-2} in a neighbourhood of zero and under the assumption $n\Delta^2 \leq 1$.
- However, if the Lévy measure behaves as $M|x|^{-(1+\alpha)}$ with $\alpha \in (1,2)$, then the rate of convergence deteriorates. The rate can be explicitly computed and it depends on α.

Gaussian approximation of the small jumps

Set $\sigma^2(\varepsilon) := \int_{|x|\leq\varepsilon} x^2\nu(dx)$. Then, by a result of Gnedenko and Kolmogorov,

$$\frac{X_t^S(\varepsilon)}{\sqrt{t\sigma^2(\varepsilon)}} \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \mathcal{N}(0,1).$$

B.V. Gnedenko and A.N. Kolmogorov, Limit distributions for sums of independent random variables. Addison-Wesley, 1954.

Gaussian approximation of the small jumps

Set $\sigma^2(\varepsilon) := \int_{|x|\leq\varepsilon} x^2\nu(dx)$. Then, by a result of Gnedenko and Kolmogorov,

$$\frac{X_t^S(\varepsilon)}{\sqrt{t\sigma^2(\varepsilon)}} \xrightarrow[\varepsilon\to 0]{\mathscr{L}} \mathcal{N}(0,1).$$

- Convergence in law is a weak notion of convergence. Is it possible to show the convergence for a meaningful metric?
- If so, what is the exact order for the rate of convergence?

B.V. Gnedenko and A.N. Kolmogorov, Limit distributions for sums of independent random variables. Addison-Wesley, 1954.

Gaussian approximation for the small jumps in Wasserstein distance

Theorem (M., Reiß)

Let $\varepsilon \in (0,1]$ *and* X *be any Lévy process with Lévy measure* ν. *For any* $p \in [1,2]$, *there exists a positive constant* C, *such that*

$$\mathscr{W}_p\Big(\mathscr{L}(X_t^S(\varepsilon)), \mathcal{N}\big(0, t\sigma^2(\varepsilon)\big)\Big) \leq C \min\left(\sqrt{t}\sigma(\varepsilon), \left(\frac{\int_{-\varepsilon}^{\varepsilon} |x|^{p+2}\nu(dx)}{\sigma^2(\varepsilon)}\right)^{1/p}\right)$$
$$\leq C \min\Big(\sqrt{t}\sigma(\varepsilon), \varepsilon\Big).$$

In particular, for $p = 1$ *the bound is* $\min(2\sqrt{t}\sigma(\varepsilon), \frac{\varepsilon}{2})$.

E. Mariucci, M. Reiß, Wasserstein and total variation distance between marginals of Lévy processes, Electronic Journal of Statistics, 2018.

Gaussian approximation for the small jumps in total variation

Theorem (Carpentier, Duval, M.)

Let $\varepsilon \in (0,1]$ *and* X *be a pure jump Lévy process with Lévy measure* ν. *Set* $\nu_\varepsilon := \nu_{|_{[-\varepsilon,\varepsilon]}}$, $\mu_3(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} x^3 \nu_\varepsilon(dx)$ *and* $\mu_4(\varepsilon) = \int_{-\varepsilon}^{\varepsilon} x^4 \nu_\varepsilon(dx)$. *Suppose that*

$$\frac{c_-}{|x|^{1+\alpha}} \le \frac{\nu_\varepsilon(dx)}{dx} \le \frac{c_+}{|x|^{1+\alpha}}, \quad \text{for some } c_+, c_- > 0, \ \alpha \in (0,2).$$

Then, there exists a constant $C > 0$ *such that for all* $n \ge 1$ *and* $\Delta > 0$ *with* $n\Delta \ge 1$ *it holds:*

$$\left\| \mathscr{L}(X_\Delta^S)^{\otimes n} - \mathcal{N}(0,\Delta(\sigma^2(\varepsilon))^{\otimes n} \right\|_{TV} \le C\left(\sqrt{\frac{n\mu_4^2(\varepsilon)}{\Delta^2(\sigma^2(\varepsilon))^4} + \frac{n\mu_3^2(\varepsilon)}{\Delta(\sigma^2(\varepsilon))^3}} + \frac{1}{n} \right).$$

A. Carpentier, C. Duval, E. Mariucci, Total variation distance for discretely observed Lévy processes: a Gaussian approximation of the small jumps, arXiv:1810.02998.

Conclusions:

- We consider the problem of the nonparametric estimation of the Lévy density from the high-frequency observations of one trajectory of a Lévy process.
- We prove the asymptotic equivalence between the Lévy density estimation problem and a Gaussian white noise model.
- We propose an estimator of the Lévy density based on a compound Poisson approximation. It achieves (nearly)-optimal minimax rates for Lévy densities in Besov balls that are bounded by $|x|^{-2}$ in a neighbourhood of zero. If the Lévy density behaves as $|x|^{-(1+\alpha)}$ for some $\alpha \in (1,2)$ in a neighbourhood of zero, then then the rates are slower.
- We provide sharp bounds for a Gaussian approximation of the small jumps in Wasserstein and total variation distance.
- The problem of including a Gaussian approximation for the small jumps in the estimation procedure and studying its statistical properties is still open.

Conclusions:

- We consider the problem of the nonparametric estimation of the Lévy density from the high-frequency observations of one trajectory of a Lévy process.
- We prove the asymptotic equivalence between the Lévy density estimation problem and a Gaussian white noise model.
- We propose an estimator of the Lévy density based on a compound Poisson approximation. It achieves (nearly)-optimal minimax rates for Lévy densities in Besov balls that are bounded by $|x|^{-2}$ in a neighbourhood of zero. If the Lévy density behaves as $|x|^{-(1+\alpha)}$ for some $\alpha \in (1,2)$ in a neighbourhood of zero, then then the rates are slower.
- We provide sharp bounds for a Gaussian approximation of the small jumps in Wasserstein and total variation distance.
- The problem of including a Gaussian approximation for the small jumps in the estimation procedure and studying its statistical properties is still open.

Thank you for your attention!