Adaptive estimation methods for pure jump Lévy processes in high frequency setting

Fabienne Comte*
MAP5, UMR CNRS 8145, University Paris Descartes, France –
fabienne.comte@parisdescartes.fr
Valentine Genon-Catalot
MAP5, UMR CNRS 8145, University Paris Descartes, France –
valentine.genon-catalot@parisdescartes.fr

This work is concerned with nonparametric estimation of the Lévy density of a Lévy process. The sample path is observed at n equispaced instants with sampling interval Δ. We develop several nonparametric adaptive methods of estimation based on deconvolution, projection and kernel. The asymptotic framework is: n tends to infinity, $\Delta = \Delta_n$ tends to 0 while $n\Delta_n$ tends to infinity (high frequency). Bounds for the $\mathbb{L}^2$-risk of several types of adaptive estimators are given in the pure jump case: to that aim, procedures including cutoff, bandwidth, or model selection are described, in the different contexts of Fourier, kernel or projection methods. Rates of convergence are discussed. A specific method for estimating the jump density of compound Poisson processes is presented. Examples and simulation results illustrate the performance of estimators. The generalization to the setting of processes including a drift or a Gaussian component is discussed.

Keywords. Adaptive estimation, Lévy process, non parametric methods, high frequency data.

1. Introduction

The aim of this contribution is to present statistical adaptive methods of estimation of the Lévy measure of a Lévy process, *i.e.* a continuous time process with stationary independent increments whose sample paths are right-continuous with left-hand limits. We refer to Bertoin (1996) or Sato (1999) for a detailed probabilistic study of these processes. In what follows, we assume that the process is real-valued, discretely observed at equispaced instants and inference is based on a sample of n observations.

The distribution of a Lévy process is usually specified by its characteristic triple, the drift, the Gaussian component and the Lévy measure rather than by the distribution of its independent increments. Indeed, the distributions of increments often have no closed form formula. This is why statistical references have increasingly focused on nonparametric methods. This presentation relies on Comte and Genon-Catalot (2014).

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

Adaptive nonparametric methods have been developed for density estimation from *i.i.d.* observations: see Donoho et al. (1996) for wavelet thresholding methods, Barron et al. (1999) or Massart (2007) for model selection and contrast penalization methods or Goldenshluger and

Lepski (2011) for data-driven bandwidth selection in kernel estimation. In the present work, we have adapted some of these approaches for estimating the Lévy density.

For *i.i.d.* data contaminated with additive noise, specific methods have been introduced, based on Fourier inversion and called deconvolution methods. The estimation of the Lévy density for Lévy processes relies on the explicit form of the characteristic function and thus takes inspiration in the deconvolution methods. The nonparametric estimation of the Lévy density has been studied for a continuous time observation of the sample path on a time interval $[0, T]$ with T tending to infinity (Figueroa-López and Houdré (2006)) or for discrete time observations. Several adaptive estimation methods have been considered in high frequency data setting and may be developed here: deconvolution with cut-off selection, contrast penalization, see our works presented in Comte and Genon-Catalot (2014), and also Figueroa-López (2009), Ueltzhöfer and Klüppelberg (2011) and adaptive kernels (see Section 3 here, and also Bec and Lacour (2014)).

We only deal with upper risk bounds, but to check the optimality of the estimators, lower bounds are needed: they are provided, in the high frequency setting by Figueroa-López (2009), Bec and Lacour (2014), and in the low frequency setting by Belomestny and Reiß (2006), Neumann and Reiß (2009), Kappus and Reiß (2010), Kappus (2014). Lower bound in the specific case of decompounding is obtained in Duval (2013).

All proposed methods rely on assumptions on the characteristic triplet of the Lévy process which are most often unknown. This is obviously the general situation for statistical inference. Nevertheless, model assumptions should be checked if possible.

2. Framework

The Lévy process is denoted by (L_t) and the observations are $(L_{k\Delta}, k = 1, \dots, n)$ where Δ is the sampling interval. The statistical procedure is based on the *i.i.d.* increments

$$Z_k^{\Delta} = L_{k\Delta} - L_{(k-1)\Delta}. \tag{1}$$

We assume that, as n tends to infinity,

$$\Delta = \Delta_n \to 0, \quad \text{and} \quad n\Delta_n \to +\infty. \tag{2}$$

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

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

where the characteristic exponent is given by

$$\psi(u) = iu\tilde{b} - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}} \left(e^{iux} - 1 - iux 1_{|x|\le 1}\right) n(x)dx, \tag{3}$$

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

$$\int_{\mathbb{R}} (x^2 \wedge 1) n(x) dx < +\infty. \tag{4}$$

Thus, $(Z_k, k = 1, \dots, n)$ is an *i.i.d.* sample with characteristic function φ_Δ. The core of this paper lies in obtaining adaptive data-driven estimation methods for the Lévy density $n(.)$. The estimation of the other parameters $\tilde{b}, \sigma^2$ may be also addressed under different sets of assumptions on the Lévy process, but we will not develop it here.

Our approaches are inspired from density estimation methods, whether directly or by deconvolution. However, for Lévy density estimation, we have to take into account the fact that the function $n(.)$ is not integrable near 0. This is why we focus on estimation of a modified Lévy density of the form $x^j n(x)$, $j = 1, 2, 3$ depending on whether $x^j n(x)$ belongs to $\mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R})$. The estimation of $n(.)$ on a compact set separated from 0 can be deduced. We use the following notations for the functions to be estimated:

$$(5) \qquad g(x) = x\, n(x), \quad \ell(x) = x^2\, n(x), \quad p(x) = x^3\, n(x).$$

Integrability around 0 and around infinity for the Lévy density correspond to different properties of the process. Around 0, integrability of $\ell(x)$ (hence of $p(x)$) is the minimal assumption for Lévy densities. The stronger constraint of integrability of g near 0 means a low activity of jumps with a Blumenthal-Getoor index smaller than 1. Around infinity, integrability of g, ℓ, p is a moment assumption for the random variable L_t.

For sake of conciseness, we present here only the estimation of g corresponding to the pure jump case, under the assumption:

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

The estimation of ℓ would require the assumption $\int_{\mathbb{R}} x^2 n(x)dx < \infty, \;\; \sigma^2 = 0$, and the estimation of p, the assumption $\int_{\mathbb{R}} |x|^3 n(x)dx < \infty$.

We also need the following assumption where the value of l will be given in statements.

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

The following proposition relates the moments of $Z_1 = L_\Delta$ under (H2-(l)) to the integrals $m_l = \int_{\mathbb{R}} x^l n(x)dx$.

Proposition 2.1. *Assume (H1-g) and (H2-(l)) with $l \geq 2$. Then,* $\mathbb{E}(Z_1) = \Delta m_1$, $\mathbb{E}(Z_1^2) = \Delta m_2 + \Delta^2 m_1^2$, *and more generally, for* $2 \leq q \leq l$,

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

Under (H1-g), the process (L_t) has finite variation on compact sets, is of pure jump type, with no drift component. Formula (3) simplifies into

$$(6) \qquad \psi(u) = \int_{\mathbb{R}} \left(e^{iux} - 1\,\right) n(x)dx.$$

The distribution of (L_t) is therefore completely specified by the knowledge of $n(.)$ which describes the jumps behavior.

We consider now a Lévy process (L_t) discretely observed with sampling interval Δ under the asymptotic framework (2) and assume that (H1-g) holds and that the characteristic exponent is given by (6). For the estimation of $g(x) = xn(x)$, (H1-g), (H2-(l)) for an integer l to be precised in each proposition or theorem and the following additional assumptions are required.

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

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

Assumptions (H1-g) and (H2-(l)) are moment assumptions for the *i.i.d.* observed random variables $(Z_k = L_{k\Delta} - L_{(k-1)\Delta}, k = 1, \dots, n)$. Under (H1-g), (H2-($l$)) for $l > 1$ implies (H2-(k)) for $k \leq l$. Noting that $\|g\|_1^2 := (\int |g(x)|dx)^2 \leq \int (1+|x|)^2 g^2(x)dx \int dx/(1+|x|)^2$, we see that (H3-g)-(H4-g) imply (H1-g).

3. Kernel estimators

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

$$\hat{g}_h(x) = K_h \star \hat{\mu}_n(x) = \frac{1}{n\Delta}\sum_{k=1}^{n} Z_k K_h(x - Z_k). \tag{7}$$

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

Definition 3.1. *A kernel K is said to be of order l if functions $u \mapsto u^j K(u), j = 0, 1, ..., l$ are integrable and satisfy $\int u^j K(u)du = 0$, $\forall j \in \{1, ..., l\}$.*

The assumptions on K are the following.

- (Ker[1, α]) K is a kernel of order $l = \lfloor \alpha \rfloor$ and $\int |x|^\alpha |K(x)|dx < +\infty$.
- (Ker[2]) $\|K\| := (\int K^2(u)du)^{1/2} < +\infty$.
- (Ker[3]) $K^* \in \mathbb{L}^1$.

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

Remark 3.1. *A kernel of order l can be built following Kerkyacharian et al. (2001).*

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

Definition 3.2. *(Nikol'ski class) Let $\alpha > 0$, $l = \lfloor \alpha \rfloor$ and $L > 0$. The Nikol'ski class $N(\alpha, L)$ on $\mathbb{R}$ is the set of all functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ such that derivatives $f^{(j)}$ for $j = 1, \ldots, l$ exist and $f^{(l)}$ verifies:*

$$\left(\int |f^{(l)}(x+t) - f^{(l)}(x)|^2 dx\right)^{1/2} \leq L|t|^{\alpha - l}, \forall t \in \mathbb{R}. \tag{8}$$

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

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

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

The bias needs further decomposition and involves the usual bias of the kernel method, $b_{h,1}(x) = K_h \star g(x) - g(x)$, and the bias resulting from the approximation of $\varphi_\Delta(u)$ by 1, $b_{h,2}(x) = \mathbb{E}[\hat{g}_h(x)] - K_h \star g(x)$, where $\star$ denotes the convolution product.

Proposition 3.1. *Under (Ker[1, α]) to (Ker[3]), (H1-g), (H2-(2)), (H3-g) and if $\int v^2|g^*(v)|^2 dv := A < +\infty$, we have*

$$\mathbb{E}(\|\hat{g}_h - g\|^2) \leq 2\|g - g \star K_h\|^2 + \frac{\|K\|^2 \mathbb{E}(Z_1^2/\Delta)}{nh\Delta} + \frac{A\|K\|_1^2 \|g\|_1^2}{\pi}\Delta^2. \tag{9}$$

If in addition $g \in N(\alpha, L)$, then $\|g - g \star K_h\|^2 \leq c_1 h^{2\alpha}$ with c_1 a constant depending on L.

We set $h = h_n$ with $h_n \to 0$ and $nh_n \to +\infty$. Recall that $\Delta = \Delta_n$ is such that $\lim_{n\to+\infty} \Delta_n = 0$. Consequently, $1/nh$ is negligible compared to $1/nh\Delta$. To obtain the optimal convergence rate based on the first two terms of (9), a constraint on Δ is necessary. We impose $\Delta^2 \leq 1/(nh\Delta)$, equivalently

$$\Delta^3 \leq \frac{1}{nh}. \tag{10}$$

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

Proposition 3.2. *Under the assumptions of Proposition 3.1 and under condition (10), the choice $h_{opt} \propto (n\Delta)^{-\frac{1}{2\alpha+1}}$ minimizes the risk bound (9) and gives*

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

4. Data-driven choice of the bandwidth and adaptive estimator

Now, α being unknown, we must select the bandwidth by a data-driven criterion. For this, adequate estimators of the dominating risk bound terms (see (9)) must be found. Following ideas given in Goldenshluger and Lepski (2011) for density estimation, we set:

$$V(h) = \kappa \|K\|_1^2 \|K\|^2 \frac{\mathbb{E}(Z_1^2/\Delta)}{nh\Delta}, \tag{11}$$

where κ is a numerical constant that will be precised below. Note that $V(h)$ is proportional to the bound of $\int \mathrm{Var}[\hat{g}_h(x)]dx$. In the above definition, $V(h)$ depends on the unknown moment $\mathbb{E}Z_1^2$. Actually, this moment is replaced by the empirical mean $n^{-1}\sum_{k=1}^n Z_k^2$, a substitution which is possible, both in theory and in practice.
The estimation of the bias term relies on iterated kernel estimators. Define

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

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

$$A(h) = \sup_{h' \in \mathcal{H}} \{\|\hat{g}_{h,h'} - \hat{g}_{h'}\|^2 - V(h')\}_+, \tag{12}$$

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

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

Theorem 4.1. *Assume (Ker[1, α]), (Ker[2]), (Ker[3]), (H2-(8))-(H3-g)-(H4-g), and $\int v^2|g^*(v)|^2dv := A < +\infty$. Assume moreover that $\mathcal{H}$ is such that $M = \mathrm{card}(\mathcal{H}) \le n\Delta$, $\forall h \in \mathcal{H}, h \ge 1/(n\Delta)$ and*

$$\forall C > 0,\ \exists \Sigma(C) < +\infty \quad \textit{such that} \quad \sum_{h \in \mathcal{H}} h^{-1/2}\exp(-Ch^{-1/2}) \le \Sigma(C). \tag{13}$$

Then we have

$$\mathbb{E}[\|g - \hat{g}_{\hat{h}}\|^2] \le c_1 \inf_{h \in \mathcal{H}} \left\{\|g - g \star K_h\|^2 + V(h)\right\} + c_2\Delta^2 + c_3 \frac{\log^2(n\Delta)}{n\Delta},$$

where c_1, c_2, c_3 are constants depending on $\|g\|, \|K\|_1$, $\|K\|, \mathbb{E}(Z_1^2/\Delta), \mathbb{E}(Z_1^4/\Delta)$, M_2 of (H4-g) and on condition (13).

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

Remark 4.1. *The infimum in the bound of Theorem 4.1 means that the estimator $\hat{g}_{\hat{h}}$ automatically reaches the optimal rate stated in Proposition 3.2.*

5. Concluding remarks

We presented above adaptive kernel estimators for $g(x) = xn(x)$ when considering a pure jump Lévy process. In Comte and Genon-Catalot (2014), two ways of extensions are considered, for both the methods (of Fourier types, or using model selection for projection estimators) and the functions under estimation ($\ell(x) = x^2 n(x)$, $p(x) = x^3 n(x)$ in presence of drift or Gaussian component).

References

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413.

Bec, M. and Lacour, C. (2014). Adaptive kernel estimation of the lévy density. *Stat. Inference Stoch. Process.*, page To appear.

Belomestny, D. and Reiß, M. (2006). Spectral calibration of exponential Lévy models. *Finance Stoch.*, 10(4):449–474.

Bertoin, J. (1996). *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge.

Comte, F. and Genon-Catalot (2014). Adaptive estimation for Lévy processes. In *Lévy Matter IV*, pages 77–177. Springer, New York.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539.

Duval, C. (2013). Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.*, 123(11):3963–3986.

Figueroa-López, J. E. (2009). Nonparametric estimation of Lévy models based on discrete-sampling. In *Optimality*, volume 57 of *IMS Lecture Notes Monogr. Ser.*, pages 117–146. Inst. Math. Statist., Beachwood, OH.

Figueroa-López, J. E. and Houdré, C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. In *High dimensional probability*, volume 51 of *IMS Lecture Notes Monogr. Ser.*, pages 96–116. Inst. Math. Statist., Beachwood, OH.

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3):1608–1632.

Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Process. Appl.*, 124(1):730–758.

Kappus, J. and Reiß, M. (2010). Estimation of the characteristics of a Lévy process observed at arbitrary frequency. *Stat. Neerl.*, 64(3):314–328.

Kerkyacharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2):137–170.

Massart, P. (2007). *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

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

Sato, K.-i. (1999). *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author.

Ueltzhöfer, F. A. J. and Klüppelberg, C. (2011). An oracle inequality for penalised projection estimation of Lévy densities from high-frequency observations. *J. Nonparametr. Stat.*, 23(4):967–989.