NONPARAMETRIC DENSITY ESTIMATION IN COMPOUND POISSON PROCESS USING CONVOLUTION POWER ESTIMATORS.

F. COMTE[1], C. DUVAL[2], AND V. GENON-CATALOT[1]

ABSTRACT. Consider a compound Poisson process which is discretely observed with sampling interval Δ until exactly n nonzero increments are obtained. The jump density and the intensity of the Poisson process are unknown. In this paper, we build and study parametric estimators of appropriate functions of the intensity, and an adaptive nonparametric estimator of the jump size density. The latter estimation method relies on nonparametric estimators of m-th convolution powers density. The L^2-risk of the adaptive estimator achieves the optimal rate in the minimax sense over Sobolev balls. Numerical simulation results on various jump densities enlight the good performances of the proposed estimator.

Keywords. Convolution. Compound Poisson process. Inverse problem. Nonparametric estimation. Parameter estimation.
AMS Classification. 62G07 - 60G51 - 62F12.

1. Introduction

Compound Poisson processes are widely used in practice especially in queuing and insurance theory (see *e.g.* Embrechts *et al.*, 1997 and refererences therein, Katz (2002) or Scalas (2006)). Let $(X_t, t \geq 0)$ be a compound Poisson process, given by

$$X_t = \sum_{i=1}^{N_t} \xi_j, \tag{1}$$

where $(\xi_j, j \geq 1)$ is a sequence of *i.i.d.* real valued random variables with density f, (N_t) is a Poisson process with intensity $c > 0$, independent of the sequence $(\xi_j, j \geq 1)$. The density f and the intensity c are unknown. In this paper, we are interested in adaptive nonparametric estimation of f from discrete observations $(X_{j\Delta}, j \geq 0)$ of the sample path with sampling interval Δ.

Compound Poisson processes have independent and stationary increments. They are a special case of Lévy processes with integrable Lévy density equal to $cf(.)$. It is therefore natural to base the estimation procedure for f on the *i.i.d.* increments $(X_{j\Delta} - X_{(j-1)\Delta}, j \geq 1)$. If c is known, the nonparametric estimation of f is equivalent to the nonparametric estimation of the Lévy density of a pure jump Lévy process with integrable Lévy measure. Several papers on the subject are available, see Basawa and Brockwell (1982), Buchmann

[1] Université Paris Descartes, MAP5, UMR CNRS 8145. emails: fabienne.comte@parisdescartes.fr and valentine.genon-catalot@parisdescartes.fr.
[2] CREST-ENSAE and CEREMADE, Université Paris Dauphine. email: Celine.Duval@ensae.fr.

(2009), Chen *et al.* (2010), Comte and Genon-Catalot (2009, 2010, 2011), Figueroa-López and Houdré (2006), Figueroa-López (2009), Gugushvili (2009, 2012), Jongbloed *et al.* (2005), Kim (1999), Neumann and Reiss (2009), Ueltzhöfer and Klüppelberg (2011), Zhao and Wu (2009).

However, specific methods for compound Poisson processes have been investigated, for instance Buchmann and Grübel (2003) first introduced decompounding methods to estimate discrete compound densities. Indeed, the common distribution of the increments is equal to

$$\mathbb{P}_{X_\Delta}(dx) = e^{-c\Delta}\delta_0(dx) + (1-e^{-c\Delta})g_\Delta(x)dx, \tag{2}$$

where δ_0 is the Dirac mass at 0, g_Δ is the conditional density of X_Δ given that $X_\Delta \neq 0$:

$$g_\Delta = \sum_{m\geq 1} \frac{e^{-c\Delta}}{1-e^{-c\Delta}}\frac{(c\Delta)^m}{m!} f^{\star\, m}, \tag{3}$$

and $f^{\star\, m}$ denotes the m-th convolution power of f. Thus, null increments provide no information on the density f. Relying on this fact, van Es *et al.* (2007) assume that the sample path X_t is discretely observed until exactly n increments are nonzero. Such observations can be described as follows. Let

$$S_1 = \inf\{j \geq 1, X_{j\Delta} - X_{(j-1)\Delta} \neq 0\}, \quad S_i = \inf\{j > S_{i-1}, X_{j\Delta} - X_{(j-1)\Delta} \neq 0\}, i \geq 2, \tag{4}$$

and let

$$Z_i = X_{S_i\Delta} - X_{(S_i-1)\Delta}. \tag{5}$$

Assume that the $X_{j\Delta}$'s are observed for $j \leq S_n$. Thus, $(S_i, Z_i), i = 1, \dots, n$ are observed and $Z_1, \dots, Z_n$ is a n-sample of the conditional distribution of X_Δ given that $X_\Delta \neq 0$ which has density g_Δ (see Proposition 2.1). The problem then is to deduce an estimator of f from an *i.i.d.* sample of g_Δ.

Under the assumption that the intensity c is known and for $\Delta = 1$ (low frequency data), van Es *et al.* (2007) build a nonparametric kernel estimator of f exploiting the relationship between the characteristic function of f and the characteristic function of g_Δ. In Duval (2012a), a different estimation method is considered. Duval (2012a) remarks that the operator $f \to g_\Delta := P_\Delta f$ can be explicitly inverted, actually using the relationship pointed out by van Es *et al.* (2007). So, $f = P_\Delta^{-1} g_\Delta$. Provided that $c\Delta < \log 2$, the inverse operator P_Δ^{-1} admits a series development given by (see Duval, 2012a, chap. 3, Lemma 1):

$$g \mapsto P_\Delta^{-1}(g) = \sum_{m\geq 1} \frac{(-1)^{m+1}}{m}\frac{(e^{c\Delta}-1)^m}{c\Delta} g^{\star m}. \tag{6}$$

Consequently, truncating the above development and keeping $K+1$ terms, an approximation of the inverse operator is obtained which suggests to approximate f by:

$$f \simeq \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m}\frac{(e^{c\Delta}-1)^m}{c\Delta} g_\Delta^{\star m}. \tag{7}$$

The approximation is valid for small Δ. Afterwards, an estimator of f is built replacing, for $m = 1, \ldots, K+1$, $(e^{c\Delta}-1)^m/c\Delta$ by a consistent estimator and $g_\Delta^{\star m}$ by a nonparametric estimator based on the observations $(Z_j, j = 1, \ldots, n)$. This is not quite simple as $g_\Delta^{\star m}$ is the density of the sum $Z_1 + \ldots, Z_m$. The estimator proposed by Duval for $g_\Delta^{\star m}$ is a wavelet threshold estimator using data composed by independent sums of m observations assuming that a deterministic number n_T of increments are observed with sampling interval Δ_T and total length time of observation $T = n_T\Delta_T$. The rate of L^p-risk of the resulting estimator of f is measured in terms of T for Δ_T tending to 0 while T tends to infinity. The usual optimal rate on Besov balls is obtained up to logarithmic factors provided that $T\Delta_T^{2K+2} = O(1)$. In Comte and Genon-Catalot (2009), the adaptive estimator of the Lévy density reaches the same rate provided that $T\Delta_T^2 = O(1)$ (without logarithmic loss and for the L^2-risk only). As soon as $K \geq 1$, Duval's estimator of f improves the result of Comte and Genon-Catalot (2009), in the case of compound Poisson processes. In Kessler (1997) a similar strategy of adding correction terms to improve parametric estimators for diffusion models is also adopted.

Nevertheless, estimating $g_\Delta^{\star m}$ by building sums of m variables from the sample $(Z_1, \ldots, Z_n)$ is heavy and numerically costly. In this paper, we build a nonparametric estimator of f relying on the approximation (7). In our approach, the difference lies in the estimation method of $g_\Delta^{\star m}$. To simplify notations, we omit the dependence on Δ for g_Δ and set

$$g := g_\Delta, \qquad g^{\star m} := g_\Delta^{\star m}. \tag{8}$$

It is well known that, from a n-sample of a density g, $\sqrt{n}$-consistent nonparametric estimators of the convolution power $g^{\star m}$, for $m \geq 2$, can be built (see *e.g.* Schick and Wefelmeyer, 2004). In a recent paper, Chesneau *et al.* (2013), propose a very simple $\sqrt{n}$-consistent estimator of the m-th convolution power $g^{\star m}$ of a density g from n *i.i.d.* random variables with density g. Of course, $m \geq 2$ is fixed and should not be too large. This is the point of view adopted here.

Let g^* denote the Fourier transform of the density g. As $(g^*)^m$ is the Fourier transform of $g^{\star m}$, Chesneau *et al.* (2013) propose to estimate $(g^*)^m$ for all $m \geq 1$, by the empirical counterpart $(\tilde{g}^*(t))^m$ with:

$$\tilde{g}^*(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itZ_j}, \tag{9}$$

leading by Fourier inversion to the estimator with cutoff ℓ,

$$\widehat{g_\ell^{\star m}}(x) = \frac{1}{2\pi}\int_{-\pi\ell}^{\pi\ell} e^{-itx}(\tilde{g}^*(t))^m dt. \tag{10}$$

Afterwards, we define:

$$\widetilde{f_{K,\ell}}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} c_m(\Delta)\widehat{g_\ell^{\star m}}(x), \quad \text{with} \quad c_m(\Delta) = \frac{(e^{c\Delta}-1)^m}{c\Delta}. \tag{11}$$

As c is unknown, this is not an estimator of f. To get an estimator $\widehat{f_{K,\ell}}(x)$ of f, we replace, for all m and Δ, $c_m(\Delta)$ by an estimator $\widehat{c_m(\Delta)}$ defined below. We study for fixed

ℓ the L^2-risk of $\widehat{f_{K,\ell}}$ and propose an adaptive (data-driven) choice $\hat{\ell}$ of ℓ. We prove that the L^2-risk of the adaptive estimator $\widehat{f_{K,\hat{\ell}}}$ attains the usual optimal rate on Sobolev balls. Moreover, the risk bounds are non asymptotic and the contribution of terms coming from the estimation of $g^{\star m}$ for $m \geq 2$ is of order $O(1/n)$. Note that the total length time of observation is, in our framework, equal to $S_n\Delta$. As n tends to infinity and Δ tends to 0, this random value is asymptotically equivalent to n. Hence, the benchmark for evaluating rates is in terms of (negative) powers of n. Note that, compared to Duval (2012a), we have no logarithmic loss in our rate, which is optimal. Indeed, the lower bound is available and our adaptive estimator is thus minimax from an asymptotic point of view.

In Section 2, we define the estimators of $c_m(\Delta), m \geq 1$ and give a bound for their L^2-risk in terms of n and Δ. In Section 3, results from Chesneau *et al.* (2013) on nonparametric estimation of m-th convolution powers of a density are recalled. Section 4 concerns the estimation of f. Our main result (Theorem 4.1) gives the L^2-risk of the adaptive estimator of f. In Section 5, the estimation method is illustrated on simulated data for various jump densities. It shows that the adaptive estimator performs well for small values of K. Section 6 gives some concluding remarks. Proofs are gathered in Section 7 and Appendix.

2. Preliminary results.

Consider a compound Poisson process given by (1) and Δ a sampling interval. Then we can prove the following result.

Proposition 2.1. *Let $S_0 = 0$ and $S_i, Z_i, i \geq 1$ be given by (4)-(5). We have, for all $i \geq 1$, $\mathbb{P}(S_i < +\infty) = 1$, $(S_i - S_{i-1}, Z_i), i \geq 1$ are independent and identically distributed random couples. For $k \geq 1$,*

$$\mathbb{P}(S_1 = k, Z_1 \leq x) = e^{-c(k-1)\Delta}(1 - e^{-c\Delta})\mathbb{P}(X_\Delta \leq x | X_\Delta \neq 0).$$

Consequently, S_1 and Z_1 are independent, the distribution of Z_1 is equal to the conditional distribution of X_Δ given $X_\Delta \neq 0$, S_1 has geometric distribution with parameter $1 - e^{-c\Delta}$. Moreover, the random variables $(S_1, Z_1, \ldots, S_i - S_{i-1}, Z_i, \ldots, S_n - S_{n-1}, Z_n)$ are independent.

Let us now study the estimation of $c_m(\Delta)$. For this, we use $(S_1, \ldots, S_n)$ which are independent of the sample $(Z_1, \ldots, Z_n)$.

Proposition 2.2. *Assume that $c \in [c_0, c_1]$ with $c_0 > 0$ and $c_1\Delta \leq \log(2)/2$. For $m \geq 1$, let*

$$H_m(\xi) = \frac{1}{(\xi - 1)^m \, \log \frac{\xi}{\xi - 1}}, \tag{12}$$

and define

$$\widehat{c_m(\Delta)} = H_m(S_n/n) \, 1_{\left\{1 + \frac{1}{e^{2c_1\Delta} - 1} \leq \frac{S_n}{n} \leq 1 + \frac{1}{e^{c_0/(2\Delta)} - 1}\right\}}. \tag{13}$$

Then,

$$\mathbb{E}\left(\widehat{c_m(\Delta)} - c_m(\Delta)\right)^2 \leq C_m \frac{\Delta^{2(m-1)}}{n}, \tag{14}$$

where C_m *has an explicit expression as a function of* c_0, c_1 *and* m.

We remark that the indicator in the definition of $\widehat{c_m(\Delta)}$ implies that the estimator is set to zero on the complement of the set $\left\{1 + \frac{1}{e^{2c_1\Delta}-1} \leq \frac{S_n}{n} \leq 1 + \frac{1}{e^{c_0/(2\Delta)}-1}\right\}$, but it is shown in the proof of Proposition 2.2 that this complement has small probability.

Note that the bound in (14) is non asymptotic and the exact value of C_m can be deduced from the proof of Proposition 2.2.

3. Estimation of the m-th convolution power $g^{\star m}$ of a density g from a n-sample of g.

We recall now results proved in Chesneau et al. (2013). An important point is that estimators of the m-th convolution power $g^{\star m}$ with L^2-risk of order $1/n$ can be built. Consider an *i.i.d.* sample of variables $Z_1, \dots, Z_n$ with density g and characteristic function g^*, the Fourier transform of g. Using the standard estimator $\tilde{g}^*$ of g^* defined by (9), Chesneau et al. (2013) propose the estimator of $g^{\star m}$ given by (10). The following bounds for this estimator are proved in Chesneau et al. (2013):

Proposition 3.1. *For* $m \geq 2$ *and all* t,

$$\mathbb{E}(|\widehat{(g^*)^m}(t) - (g^*)^m(t)|^2) \leq \mathcal{E}_m \left(\frac{1}{n^m} + \frac{|g^*(t)|^2}{n}\right) \tag{15}$$

where $\mathcal{E}_m$ *is a constant which does not depend on* n *nor on* g*, increasing with* m *and* $\widehat{(g^*)^m}(t) = (\tilde{g}^*(t))^m$ *(see (9)). Consequently,*

$$\mathbb{E}(\|\widehat{g_\ell^{\star m}} - g^{\star m}\|^2) \leq \frac{1}{2\pi}\int_{|t|\geq \pi\ell} |(g^{\star m})^*(t)|^2 dt + \mathcal{E}_m\left(\frac{\ell}{n^m} + \frac{\|g\|^2}{n}\right).$$

Let us introduce the Sobolev ball

$$\mathcal{S}(\alpha, R) = \{f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R}), \int (1+x^2)^\alpha |f^*(x)|^2 dx \leq R\}.$$

If $g^{\star m}$ belongs to $\mathcal{S}(\alpha_m, R_m)$, the L^2-risk bound becomes

$$\mathbb{E}(\|\widehat{g_\ell^{\star m}} - g^{\star m}\|^2) \leq R_m \ell^{-2\alpha_m} + \mathcal{E}_m\left(\frac{\ell}{n^m} + \frac{\|g\|^2}{n}\right).$$

Choosing a trade-off bandwidth $\ell_{opt} = Cn^{m/(2\alpha_m+1)}$, we get a risk bound on $\mathbb{E}(\|\widehat{g_{\ell_{opt}}^{\star m}} - g\|^2)$ of order $\max(n^{-2m\alpha_m/(2\alpha_m+1)}, n^{-1})$. This allows to obtain a rate of order $1/n$ whenever $2m\alpha_m/(2\alpha_m+1) \geq 1$ i.e. $2\alpha_m(m-1) \geq 1$. This occurs for instance if $m \geq 2$ and $\alpha_m \geq 1/2$.

4. Estimation of f.

Let us first give the links between Sobolev regularities of f, g and $g^{\star m}$ with $g = g_\Delta$. Below, for any function $h \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$, we denote by h_ℓ the function defined by $h_\ell^* = h^* 1_{[-\pi\ell, \pi\ell]}$.

Proposition 4.1.
Let the density f belong to $\mathcal{S}(\alpha, R)$. Then g defined by (3) and (8) belongs to $\mathcal{S}(\alpha, R)$ and $g^{\star m} \in \mathcal{S}(m\alpha, R_m)$ for some constant R_m. In particular,

$$\|g\| \leq \|f\|.$$

We assume now that $c \in [c_0, c_1]$ with $c_1\Delta \leq \log 2/2$ and consider the estimator $\widehat{f_{K,\ell}}$ given by

$$\widehat{f_{K,\ell}}(x) = \sum_{m=1}^{K+1} \frac{(-1)^{m+1}}{m} \widehat{c_m(\Delta)} \widehat{g_\ell^{\star m}}(x). \tag{16}$$

where $c_m(\Delta)$ is defined in (11) and $\widehat{c_m(\Delta)}$ is the estimator of $c_m(\Delta)$ given in (13).

Proposition 4.2. *Assume that $c \in [c_0, c_1]$ with $c_0 > 0$ and $c_1\Delta \leq \log 2/2$. Then the estimator $\widehat{f_{K,\ell}}$ is such that*

$$\mathbb{E}(\|\widehat{f_{K,\ell}} - f\|^2) \leq \frac{5}{2\pi} \int_{|t|\geq \pi\ell} |f^*(t)|^2 dt + \frac{10\ell}{n} + 5A_K\Delta^{2K+2} + \frac{5B_K}{n}, \tag{17}$$

with

$$A_K = 6\frac{\|f\|^2}{(K+2)^2}(\sqrt{2}c)^{2K+2}, \tag{18}$$

(19)

$$B_K = 2(K+1)\left\{C_1(1+\|f\|^2) + \Delta^2 \sum_{m=2}^{K+1} \frac{(C_m + 2^m c^{2(m-1)})\mathcal{E}_m}{m^2} \Delta^{2(m-2)}(1+2\|f\|^2)\right\},$$

where $C_m, \mathcal{E}_m$ are the constants appearing respectively in (14) and in (15).

If $f \in \mathcal{S}(\alpha, R)$, choosing $\ell = \ell^* \propto n^{-1/(2\alpha+1)}$, inequality (17) yields

$$\mathbb{E}(\|\widehat{f_{K,\ell^*}} - f\|^2) \leq Cn^{-2\alpha/(2\alpha+1)} + 5A_K\Delta^{2K+2}. \tag{20}$$

Usually, in high frequency data for continuous time models, rates are measured in terms of the total length time of observation which is, in our framework, equal to $S_n\Delta$. Evaluating this random value as n tends to infinity, Δ tends to 0, we get that

$$S_n\Delta = \frac{S_n}{n} n\Delta \sim \frac{\Delta}{p(\Delta)} n \sim \frac{n}{c}.$$

The total length time of observation is asymptotically equivalent to n. Hence, the rate in (20) is exactly the one obtained by Duval (2012a), with no logarithmic loss.
Now, we aim at obtaining the choice of ℓ in an automatic and nonasymptotic way. For this, we propose an adaptive selection procedure.

More precisely, let

$$\hat{\ell} = \arg \min_{\ell \in \{1,2,\ldots,L_n\}} \left\{ -\|\widehat{f_{K,\ell}}\|^2 + \text{pen}(\ell) \right\}, \text{ with } \text{pen}(\ell) = \kappa \frac{\ell}{n}.$$

We can prove the following result.

Theorem 4.1. *Assume that f is bounded and $L_n \leq n$. There exists a value κ_0 such that for any κ larger than κ_0, we get,*

$$\mathbb{E}(\|\widehat{f_{K,\hat\ell}} - f\|^2) \leq 4 \min_{1\leq \ell\leq L_n} \left(\|f-f_\ell\|^2 + \mathrm{pen}(\ell)\right) + 32A_K\Delta^{2K+2} + 32\frac{B_K}{n} + \frac{C'}{n}, \tag{21}$$

where C' is a constant.

Comparing the above inequality with (17), we see that the adaptive estimator automatically realizes the best compromise between the squared bias term (first one, inside the min) and the variance term (second one, inside the min). The last two terms are standardly negligible. For the term $32A_K\Delta^{2K+2}$, either the sampling interval Δ for given K is tuned to make it negligible ($O(1/n)$) or n, Δ are given and K is chosen so that $n\Delta^{2K+2} \simeq 1$.

We now recall a lower bound derived in Duval (2012a). This lower bound is obtained in the super experiment where the compound Poisson process $(X_t, t \geq 0)$ is continuously observed over $[0, S_n\Delta]$. In that super experiment we observe (at least) n independent realizations of f and we obtain a lower bound applying classical results (see *e.g.* Tsybakov 2009).

Proposition 4.3. *We have*

$$\varliminf_{n\to\infty} \inf_{\widehat{f}} \sup_{f\in S(\alpha,R)} n^{2\alpha/(2\alpha+1)}\, \mathbb{E}(\|\widehat{f}-f\|^2) > 0 \tag{22}$$

where the infimum is taken over all estimators based on the observations $(X_t, t \leq S_n\Delta)$.

The above inequality shows that the estimator is minimax whenever $n^{-2\alpha/(2\alpha+1)}$ is larger than Δ^{2K+2}. Since $2\alpha/(2\alpha+1) \leq 1$, we take K such that $n\Delta^{2K+2} \simeq 1$.

5. Simulations

In this section we illustrate the method on simulated data. We have implemented the adaptive estimator on different examples of jump densities f, namely,

(1) A Gaussian $\mathcal{N}(0, 1)$.
(2) A Laplace $L(0, 1)$ with density $\exp(-|x|)/2$.
(3) A Gamma $\Gamma(5, 1)$.
(4) A mixture of a Gaussian and a Gamma $\frac{2}{3}\mathcal{N}(-4, 1) + \frac{1}{3}\Gamma(3, 1)$.

After preliminary experiments the constant κ is taken equal to 17.6 and the cutoff $\hat\ell$ is selected among 100 equispaced values between 0 and 10. We consider different values of Δ: $0.2, 0.5, 0.8$. For each Δ we choose K such that $n\Delta^{2K+2} \leq 1$; more precisely the corresponding values of K are $2, 5, 17$ respectively. It ensures that the estimator is minimax (see Theorem 4.1 and Proposition 4.3).

Results are given in Table 1, where 50 estimated curves are plotted on the same figure to show the small variability of the estimator. We take a sample size $n = 5000$ and an intensity $c = 0.5$, the first column gives the result for $\Delta = 0.2$ ($K = 2$), the second for $\Delta = 0.5$ ($K = 5$) and the last for $\Delta = 0.8$ ($K = 17$). On top of each graph we give the mean of selected values for $\hat\ell$ and the associated standard deviation in parenthesis evaluated

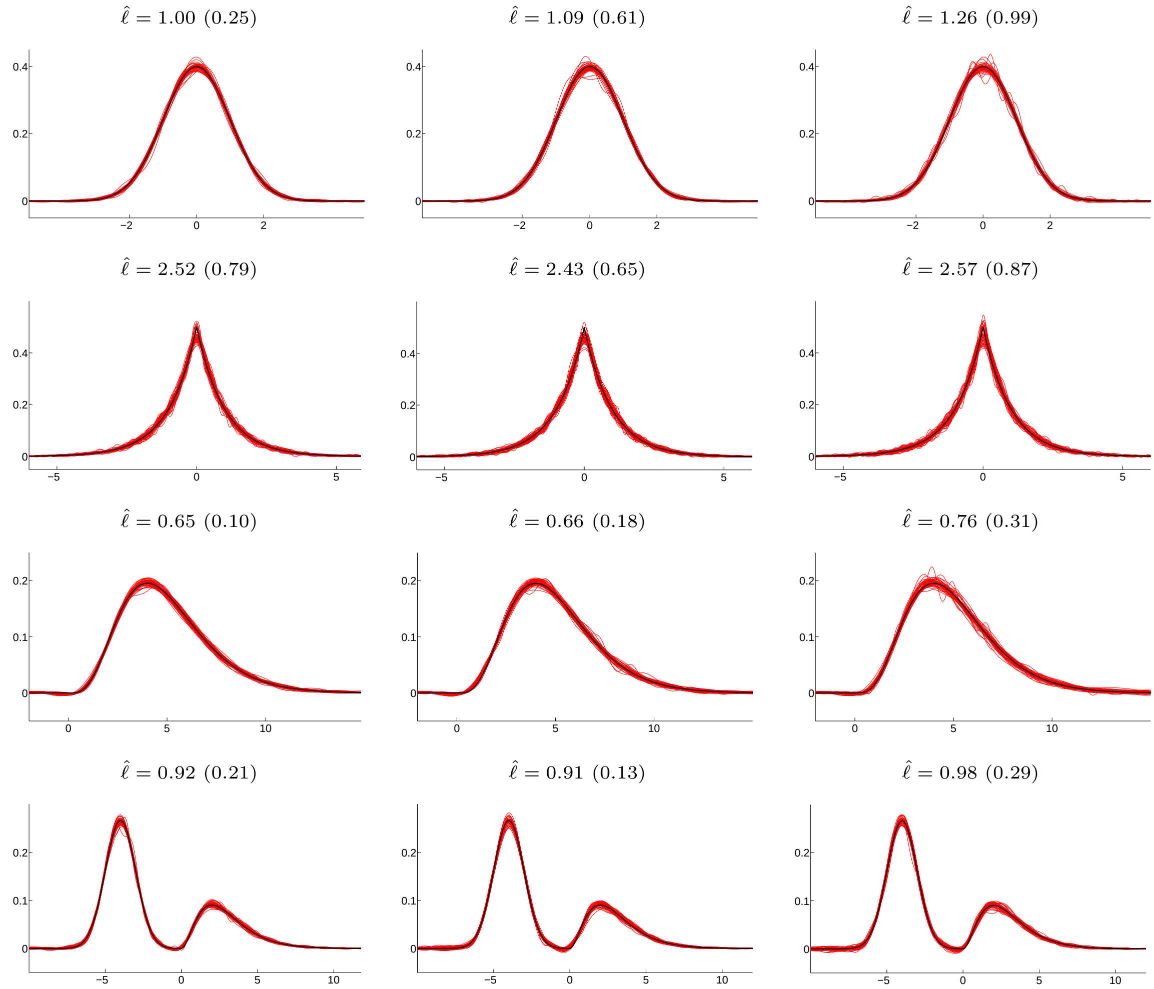

Table 1. Estimation of f for a Gaussian $\mathcal{N}(0,1)$ (first line), Laplace $L(0,1)$ second line, Gamma $\Gamma(5,1)$ (third line) and the mixture $\frac{2}{3}\mathcal{N}(-4,1)+\frac{1}{3}\Gamma(3,1)$ (fourth line) with $c=0.5$ and $n=5000$. True density (bold blak line) and 50 estimated curves (red lines), left $\Delta=0.2$ and $K=2$; middle $\Delta=0.5$ and $K=5$; right $\Delta=0.8$ and $K=17$. The value $\hat{\ell}$ is the mean over the 50 selected $\hat{\ell}$'s (with standard deviation in parenthesis).

over the fifty plots given. It appears that for each Δ the estimator well reproduces the estimated density with little variability. Increasing Δ, and therefore K, does not affect the accuracy nor the variability of the estimator.

6. Concluding remarks

In this paper, we propose a nonparametric estimator of the jump density f of a compound Poisson process. The process (X_t) is discretely observed with sampling interval Δ until exactly n nonzero increments are obtained. This provides a n-sample of the conditional distribution g_Δ of X_Δ given $X_\Delta \neq 0$. The setting is more general than in van Es *et al.* (2007), as the intensity of the Poisson process is unknown and is estimated. By inverting the operator $P_\Delta : f \mapsto P_\Delta f = g_\Delta$, we define a class of nonparametric estimators of f depending on a cutoff parameter ℓ and a truncation parameter K. For given K and small Δ, we define an adaptive choice of ℓ and prove that the resulting adaptive estimator is minimax over Sobolev balls. The estimator is easy to implement and performs well even for small K.

An interesting development would be to look for an adaptive choice of both ℓ and K by including the term $A_K \Delta^{K+2}$ in the penalty, K being searched in a finite set of integers.

Another direction, investigated by Duval (2012b) with wavelet estimators, would be an extension to renewal processes: but the lack of independence between increments makes the theoretical study much more tedious.

7. Appendix: Proofs

7.1. **Proof of Proposition 2.1.** The joint distribution of (S_1, Z_1) is elementary using that the increments $X_{j\Delta} - X_{(j-1)\Delta}$ are *i.i.d.*. The process $(X^x_{j\Delta} = x + X_{j\Delta}, j \geq 1)$ is strong Markov. We denote by $\mathbb{P}_x$ its distribution on the canonical space $\mathbb{R}^{\mathbb{N}}$, denote by $(X_j, j \geq 0)$ the canonical process of $\mathbb{R}^{\mathbb{N}}$ and by $\mathcal{F}_j = \sigma(X_k, k \leq j)$ the canonical filtration. Let $\theta : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ denote the shift operator. Consider the stopping times built on the canonical process $S_0 = 0$,

$$S_1 = \inf\{j \geq 1, X_j - X_{j-1} \neq 0\}, \quad S_i = \inf\{j > S_{i-1}, X_j - X_{j-1} \neq 0\}, i \geq 2, \tag{23}$$

and let

$$Z_i = X_{S_i} - X_{S_i - 1}. \tag{24}$$

Because the S_i's are built using the increments $(X_j - X_{j-1}, j \geq 1)$, their distributions under $\mathbb{P}_x$ is independent of the initial condition x. We have $S_i = S_{i-1} + S_1 \circ \theta_{S_{i-1}}$. The process $(X_{S_{i-1}+j} - X_{S_{i-1}} = (X_j - X_0) \circ \theta_{S_{i-1}}, j \geq 0)$ is independent of $\mathcal{F}_{S_{i-1}}$ and has distribution $\mathbb{P}_0$ and $Z_i = Z_1 \circ \theta_{S_{i-1}}$. Consequently,

$$\mathbb{E}_x(\varphi(S_i - S_{i-1})\psi(Z_i)|\mathcal{F}_{S_{i-1}}) = \mathbb{E}_0(\varphi(S_1)\psi(Z_1)).$$

By iterate conditioning, we get the result. □

7.2. **Proof of Proposition 2.2.** Let us set

$$p(\Delta) = 1 - e^{-c\Delta} = \frac{e^{c\Delta} - 1}{e^{c\Delta}}.$$

An elementary computation yields:

$$c\Delta = \log\left(\frac{x}{x-1}\right) \quad \text{with} \quad x := x(\Delta) = \frac{1}{p(\Delta)} = 1 + \frac{1}{e^{c\Delta} - 1} > 1,$$

and

$$\frac{(e^{c\Delta}-1)^m}{c\Delta} = H_m(x).$$

As the standard maximum likelihood (and unbiased) estimator of $1/p(\Delta)$ computed from the sample $(S_i - S_{i-1}, i = 1, \dots, n)$ is $S_n/n \geq 1$, we are tempted to estimate $H_m(x)$ by $H_m(S_n/n)$. This is not possible as S_n/n may be equal to 1. This is why we introduce a truncation. Set $u_0 = \Delta/(e^{c_0\Delta/2}-1)$, $u_1 = \Delta/(e^{2c_1\Delta}-1), u = \Delta/(e^{c\Delta}-1)$. Note that

$$1 + \frac{u_1}{\Delta} < x = 1 + \frac{u}{\Delta} < 1 + \frac{u_0}{\Delta}.$$

We have

$$\widehat{c_m(\Delta)} - c_m(\Delta) = H_m(S_n/n)1_{(1+\frac{u_1}{\Delta}\leq \frac{S_n}{n}\leq 1+\frac{u_0}{\Delta})} - H_m(x) = A_1 + A_2$$

with

$$A_1 = (H_m(S_n/n) - H_m(x))\ 1_{(1+\frac{u_1}{\Delta}\leq \frac{S_n}{n}\leq 1+\frac{u_0}{\Delta})}, \quad A_2 = -H_m(x)(1_{(\frac{S_n}{n}<1+\frac{u_1}{\Delta})} + 1_{(\frac{S_n}{n}>1+\frac{u_0}{\Delta})}).$$

Thus, on the set $(1+\frac{u_1}{\Delta}\leq \frac{S_n}{n}\leq 1+\frac{u_0}{\Delta})$,

$$(H_m(S_n/n) - H_m(x))^2 \leq (\frac{S_n}{n} - x)^2 \sup_{\xi\in[1+\frac{u_1}{\Delta}, 1+\frac{u_0}{\Delta}]} (H_m'(\xi))^2.$$

As

$$H_m'(\xi) = -\frac{m}{(\xi-1)^{m+1}\ \log\frac{\xi}{\xi-1}} + \frac{1}{\xi(\xi-1)^{m+1}\ \log^2\frac{\xi}{\xi-1}},$$

we have, for $\xi \in [1+\frac{u_1}{\Delta}, 1+\frac{u_0}{\Delta}]$,

$$|H_m'(\xi)| \leq \frac{2\Delta^m}{c_0 u_1^{m+1}}\left(m + \frac{2}{u_1 c_0}\right).$$

Writing that $e^{2c_1\Delta} - 1 = 2c_1\Delta e^{2sc_1\Delta}$ for $s \in (0,1)$, using that $2c_1\Delta \leq \log(2)$, we get $1/u_1 \leq 4c_1$. As

$$\mathbb{E}(\frac{S_n}{n} - x)^2 = \frac{1-p(\Delta)}{np^2(\Delta)} = \frac{e^{c\Delta}}{n(e^{c\Delta}-1)^2},$$

we get, using $e^{c\Delta} - 1 \geq c\Delta \geq c_0\Delta$:

$$\mathbb{E}A_1^2 \leq C_m' \frac{\Delta^{2(m-1)}}{n}, \text{ with } C_m' = \frac{4\sqrt{2}(4c_1)^{2(m+1)}}{c_0^4}\left(m + \frac{8c_1}{c_0}\right)^2.$$

Then, we have, setting $a_0 = u_0 - u > 0$, $a_1 = u - u_1 > 0$,

$$\begin{aligned}
\mathbb{P}\left(\frac{S_n}{n} < 1 + \frac{u_1}{\Delta}\right) + \mathbb{P}\left(\frac{S_n}{n} > 1 + \frac{u_0}{\Delta}\right) &= \mathbb{P}(\frac{\Delta}{p(\Delta)} - \Delta\frac{S_n}{n} > a_1) + \mathbb{P}(\Delta\frac{S_n}{n} - \frac{\Delta}{p(\Delta)} > a_0) \\
&\leq (\frac{1}{a_1^2} + \frac{1}{a_0^2})\frac{\Delta^2\ e^{c\Delta}}{n(e^{c\Delta}-1)^2}.
\end{aligned}$$

Thus, noting that $u_0 - u \geq 1/(2c_1)$ and $u - u_1 \geq 1/(4\sqrt{2}c_0)$,

$$\mathbb{E}A_2^2 \leq (\frac{1}{a_1^2} + \frac{1}{a_0^2})\frac{(e^{c\Delta}-1)^{2(m-1)}e^{c\Delta}}{nc^2} \leq C''_m \frac{\Delta^{2(m-1)}}{n}, \tag{25}$$

where

$$C''_m = 4\sqrt{2}\left[8c_0^2 + c_1^2\right]\frac{(4c_1)^{2(m-1)}}{c_0^2}.$$

The proof is complete with $C_m = 2(C'_m + C''_m)$. □

7.3. **Proof of Proposition 4.1.** Consider f integrable with $\|f\|_1 = \int |f|$ and square integrable such that $\int (1+x^2)^\alpha |f^*(x)|^2 dx \leq R$. Then

$$\begin{aligned}
\int (1+x^2)^\alpha |g^*(x)|^2 dx &= \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}}\right)^2 \sum_{m,k\geq 1} \frac{(c\Delta)^m}{m!}\frac{(c\Delta)^k}{k!}\int (1+x^2)^\alpha [f^*(x)]^m [f^*(-x)]^k dx \\
&\leq \left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}}\right)^2 \sum_{m,k\geq 1} \frac{(c\Delta)^m}{m!}\frac{(c\Delta)^k}{k!}\|f\|_1^{m+k-2}\int (1+x^2)^\alpha |f^*(x)|^2 dx \\
&\leq R\left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}}\right)^2 \frac{1}{\|f\|_1^2}\left(\sum_{m\geq 1}\frac{(c\Delta)^m}{m!}\|f\|_1^m\right)^2 \\
&= R\left(\frac{e^{-c\Delta}}{1-e^{-c\Delta}}\frac{\exp(c\Delta\|f\|_1)-1}{\|f\|_1}\right)^2 := R(\Delta) < +\infty
\end{aligned}$$

As f is a density, $\|f\|_1 = 1$ and $R(\Delta) = R$. This implies the announced result for g. If the density g belongs to $\mathcal{S}(\alpha, R)$, then $(1+x^2)^\alpha |g^*(x)|^2$ is continuous and integrable, thus bounded by B. Therefore $g^{\star m} \in \mathcal{S}(m\alpha, R_m)$ with $R_m = B^{m-1}R$. □

7.4. **Proof of Proposition 4.2.** Recall that $f^* = \sum_{m\geq 1}((-1)^{m+1}/m)c_m(\Delta)(g^*)^m$ (see (6)-(7)). Let f_ℓ be such that $f_\ell^* = f^*\, 1_{[-\pi\ell,\pi\ell]}$ and $f_{K,\ell}$ be such that

$$f_{K,\ell}^* = 1_{[-\pi\ell,\pi\ell]}\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}c_m(\Delta)(g^*)^m.$$

Recall that $\widetilde{f_{K,\ell}}$ (see (11)) is such that

$$(\widetilde{f_{K,\ell}})^* = 1_{[-\pi\ell,\pi\ell]}\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}c_m(\Delta)\widehat{(g^*)^m}.$$

We distinguish the first term of this development from the other ones and set

$$\widetilde{f_{K,\ell}} = \widetilde{f_{K,\ell}}^{(1)} + \widetilde{\mathcal{R}f_{K,\ell}}, \text{ with } \widetilde{f_{K,\ell}}^{(1)} = c_1(\Delta)\widehat{g_\ell^{\star 1}} = c_1(\Delta)\widehat{g_\ell}. \tag{26}$$

Analogously, with g_ℓ such that $g_\ell^* = g^*\, 1_{[-\pi\ell,\pi\ell]}$,

$$f_{K,\ell} = f_{K,\ell}^{(1)} + \mathcal{R}f_{K,\ell}, \text{ with } f_{K,\ell}^{(1)} = c_1(\Delta)g_\ell \tag{27}$$

The following decomposition of the L^2-norm holds:

$$\begin{aligned} \|f-\widehat{f_{K,\ell}}\| &\leq \|f-f_\ell\|+\|f_\ell-f_{K,\ell}\|+\|f_{K,\ell}^{(1)}-\widetilde{f_{K,\ell}}^{(1)}\| \\ &+\|\mathcal{R}f_{K,\ell}-\widetilde{\mathcal{R}f_{K,\ell}}\|+\|\widetilde{f_{K,\ell}}-\widehat{f_{K,\ell}}\|, \end{aligned}$$

which involves two bias terms and two stochastic error terms. The first bias term is the usual deconvolution bias term:

$$\|f-f_\ell\|^2=\frac{1}{2\pi}\int_{|t|\geq\pi\ell}|f^*(t)|^2dt$$

Noting that

$$f_\ell^*-f_{K,\ell}^*=1_{[-\pi\ell,\pi\ell]}\sum_{m=K+2}^{\infty}\frac{(-1)^{m+1}}{m}c_m(\Delta)(g^*)^m,$$

we get, using that $|g^*(t)|\leq 1$ and $\|g\|\leq\|f\|$ (see Proposition 4.1):

$$\begin{aligned} 2\pi\|f_\ell-f_{K,\ell}\|^2 &= \|f_\ell^*-f_{K,\ell}^*\|^2=\int_{-\pi\ell}^{\pi\ell}\left|\sum_{m=K+2}^{\infty}\frac{(-1)^{m+1}}{m}c_m(\Delta)(g^*)^m(t)\right|^2dt \\ &\leq \int_{-\pi\ell}^{\pi\ell}\left(\sum_{m\geq K+2}\frac{1}{m}c_m(\Delta)|g^*(t)|\right)^2dt\leq 2\pi\|g\|^2\left(\sum_{m\geq K+2}\frac{1}{m}c_m(\Delta)\right)^2 \\ &\leq \frac{2\pi\|f\|^2}{(c\Delta)^2(K+2)^2}\left(\frac{(e^{c\Delta}-1)^{K+2}}{2-e^{c\Delta}}\right)^2 \\ &\leq \frac{4\pi\|f\|^2(\sqrt{2}c\Delta)^{2K+2}}{((K+2)^2(2-e^{2\Delta}))^2}\leq 2\pi A_K\Delta^{2K+2}, \end{aligned} \tag{28}$$

where in the last line, we have used $1/(2-e^{c\Delta})^2\leq 1/(2-\sqrt{2})^2\leq 3$ and $e^{c\Delta}-1\leq\sqrt{2}c\Delta$ and A_K is given in (18).

To study the next term, we recall that, $\mathbb{E}(|\widehat{(g^*)(t)}-(g^*)(t)|^2)\leq 1/n$. Then we get

$$\begin{aligned} 2\pi\mathbb{E}\left(\|f_{K,\ell}^{(1)}-\widetilde{f_{K,\ell}}^{(1)}\|^2\right) &= \int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left(\left|c_1(\Delta)[\widehat{(g^*)(t)}-(g^*)(t)]\right|^2\right)dt \\ &\leq \frac{2\pi\ell[c_1(\Delta)]^2}{n}\leq\frac{4\pi\ell}{n} \end{aligned} \tag{29}$$

since $c_1(\Delta)\leq\sqrt{2}$.

Hereafter, we use inequality (15) of Proposition 3.1.

$$\begin{aligned} 2\pi\mathbb{E}\left(\|\mathcal{R}f_{K,\ell}-\widetilde{\mathcal{R}f_{K,\ell}}\|^2\right) &= \int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left(\left|\sum_{m=2}^{K+1}\frac{(-1)^{m+1}}{m}c_m(\Delta)[\widehat{(g^*)^m(t)}-(g^*)^m(t)]\right|^2\right)dt \\ &\leq \int_{-\pi\ell}^{\pi\ell}(K+1)\sum_{m=2}^{K+1}\frac{1}{m^2}[c_m(\Delta)]^2\mathbb{E}\left(|\widehat{(g^*)^m(t)}-(g^*)^m(t)|^2\right)dt \\ &\leq 2\pi K\sum_{m=2}^{K+1}\frac{\mathcal{E}_m}{m^2}[c_m(\Delta)]^2\left(\frac{\ell}{n^m}+\frac{\|g\|^2}{n}\right) \end{aligned}$$

This yields, since $c_m(\Delta)\leq(\sqrt{2})^m(c\Delta)^{m-1}$ and $\ell/n\leq 1$,

$$\mathbb{E}\left(\|\mathcal{R}f_{K,\ell}-\widetilde{\mathcal{R}f_{K,\ell}}\|^2\right)\leq\frac{D_K}{n} \tag{30}$$

with

$$D_K=K\sum_{m=2}^{K+1}\frac{2^m c^{2(m-1)}\mathcal{E}_m}{m^2}\Delta^{2(m-1)}\left(\frac{1}{n^{m-2}}+\|g\|^2\right)$$

For the last term, we use Proposition 2.2, with the fact that the estimators $\widehat{c_m(\Delta)}$ and $\widehat{(g^*)^m(t)}$ are independent, and write

$$2\pi\mathbb{E}\left(\|\widetilde{f_{K,\ell}}-\widehat{f_{K,\ell}}\|^2\right)=\int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left(\left|\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}\left(\widehat{c_m(\Delta)}-c_m(\Delta)\right)\widehat{(g^*)^m(t)}\right|^2dt\right)$$

$$\begin{aligned} &\leq 2\int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left(\left|\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}\left(\widehat{c_m(\Delta)}-c_m(\Delta)\right)[\widehat{(g^*)^m(t)}-(g^*)^m(t)]\right|^2dt\right) \\ &\quad+2\int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left(\left|\sum_{m=1}^{K+1}\frac{(-1)^{m+1}}{m}\left(\widehat{c_m(\Delta)}-c_m(\Delta)\right)(g^*)^m(t)\right|^2dt\right) \\ &\leq 2(K+1)\sum_{m=1}^{K+1}\frac{1}{m^2}\left\{\mathbb{E}\left[\left(\widehat{c_m(\Delta)}-c_m(\Delta)\right)^2\right]\int_{-\pi\ell}^{\pi\ell}\mathbb{E}\left[|\widehat{(g^*)^m(t)}-(g^*)^m(t)|^2\right]dt\right. \\ &\qquad\left.+\mathbb{E}\left[\left(\widehat{c_m(\Delta)}-c_m(\Delta)\right)^2\right]\int_{-\pi\ell}^{\pi\ell}|g^*(t)|^{2m}dt\right\} \\ &\leq 2(K+1)\left\{\frac{C_1}{n}(\frac{2\pi\ell}{n}+2\pi\|g\|^2)+\sum_{m=2}^{K+1}\frac{C_m\Delta^{2(m-1)}}{m^2}\left[\frac{\mathcal{E}_m}{n}\int_{-\pi\ell}^{\pi\ell}\left(\frac{1}{n^m}+\frac{1}{n}|g^*(t)|^2\right)dt+\frac{1}{n}\|g^*\|^2\right]\right\} \\ &\leq\frac{2\pi E_K}{n} \qquad (31) \end{aligned}$$

using that $\ell/n \leq 1$ and

$$E_K = 2(K+1)\left[C_1(1+\|g\|^2) + \sum_{m=2}^{K+1} \frac{C_m}{m^2}\Delta^{2(m-1)}\mathcal{E}_m(\frac{1}{n^{m-1}} + 2\|g\|^2)\right].$$

This ends the proof of the result with $D_K + E_K \leq B_K$ and $\|g\| \leq \|f\|$. □

7.5. **Proof of Theorem 4.1.** Consider the contrast

$$\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{f_{K,L_n}}\rangle,$$

and for $\ell = 1, \dots, L_n$, the increasing sequence of spaces

$$S_\ell = \{t \in \mathbb{L}^2 \cap \mathbb{L}^1(\mathbb{R}),\ \text{supp}(t^*) \subset [-\pi\ell, \pi\ell]\}.$$

Note that, for $\ell \leq L_n$ and $t \in S_\ell$, $\gamma_n(t) = \|t\|^2 - 2\langle t, \widehat{f_{K,\ell}}\rangle$, and

$$\arg\min_{t\in S_\ell} \gamma_n(t) = \widehat{f_{K,\ell}}, \quad \text{with} \quad \gamma_n(\widehat{f_{K,\ell}}) = -\|\widehat{f_{K,\ell}}\|^2.$$

For $\ell, \ell* \leq L_n$, $s \in S_\ell$ and $t \in S_{\ell*}$, the following decomposition holds:

$$\gamma_n(t) - \gamma_n(s) = \|t-f\|^2 - \|s-f\|^2 - 2\langle t-s, \widehat{f_{K,L_n}} - f\rangle$$

and $\langle t-s, \widehat{f_{K,L_n}} - f\rangle = \langle t-s, \widehat{f_{K,L_n}} - f_{L_n}\rangle$. By definition of the estimator,

$$\gamma_n(\widehat{f_{K,\hat\ell}}) + \text{pen}(\hat\ell) \leq \gamma_n(\widehat{f_{K,\ell}}) + \text{pen}(\ell) \leq \gamma_n(f_\ell) + \text{pen}(\ell).$$

Thus, we obtain, $\forall \ell \in \{1, \dots, L_n\}$,

$$\begin{aligned}
\|\widehat{f_{K,\hat\ell}} - f\|^2 &\leq \|f_\ell - f\|^2 + \text{pen}(\ell) + 2\langle \widehat{f_{K,\hat\ell}} - f_\ell, \widehat{f_{K,L_n}} - f_{L_n}\rangle - \text{pen}(\hat\ell) \\
&\leq \|f_\ell - f\|^2 + \text{pen}(\ell) + \frac{1}{4}\|\widehat{f_{K,\hat\ell}} - f_\ell\|^2 + 4\sup_{t\in S_\ell + S_{\hat\ell}, \|t\|=1} \langle t, \widehat{f_{K,L_n}} - f_{L_n}\rangle^2 - \text{pen}(\hat\ell) \qquad (32)
\end{aligned}$$

Then using

$$\frac{1}{4}\|\widehat{f_{K,\hat\ell}} - f_\ell\|^2 \leq \frac{1}{2}\|\widehat{f_{K,\hat\ell}} - f\|^2 + \frac{1}{2}\|f - f_\ell\|^2, \qquad (33)$$

and decompositions (26) and (27), we get

$$\langle t, \widehat{f_{K,L_n}} - f_{L_n}\rangle = \langle t, \widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}}\rangle + \langle t, \widetilde{f_{K,L_n}}^{(1)} - f^{(1)}_{K,L_n}\rangle + \langle t, \mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n}\rangle + \langle t, f_{K,L_n} - f_{L_n}\rangle.$$

By the Cauchy-Schwarz Inequality and for $\|t\| = 1$, we have

$$\begin{aligned}
\langle t, \widehat{f_{K,L_n}} - f_{L_n}\rangle^2 &\leq 4\|\widehat{f_{K,L_n}} - \widetilde{f_{K,L_n}}\|^2 + 4\|\mathcal{R}\widetilde{f_{K,L_n}} - \mathcal{R}f_{K,L_n}\|^2 \\
&\quad + 4\|f_{K,L_n} - f_{L_n}\|^2 + 4\langle t, \widetilde{f_{K,L_n}}^{(1)} - f^{(1)}_{K,L_n}\rangle^2. \qquad (34)
\end{aligned}$$

Thus, inserting (33) and (34) in (32) yields

$$\begin{aligned}\frac{1}{2}\|\widehat{f_{K,\hat\ell}}-f\|^2 \quad \le \quad & \frac{3}{2}\|f_\ell - f\|^2 + 16\|f_{K,L_n}-f_{L_n}\|^2 \\ & +16\|\widehat{f_{K,L_n}}-\widetilde{f_{K,L_n}}\|^2 + 16\|\mathcal{R}\widetilde{f_{K,L_n}}-\mathcal{R}f_{K,L_n}\|^2 + \mathrm{pen}(\ell) \\ & +16 \sup_{t\in S_{\ell\vee\hat\ell}, \|t\|=1} \langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)}\rangle^2 - \mathrm{pen}(\hat\ell)\end{aligned}$$

Here, the bounds of Proposition 4.2 can be applied. Indeed (28), (30) and (31) are uniform with respect to ℓ and imply

$$\|f_{K,L_n}-f_{L_n}\|^2 \le A_K\Delta^{2(K+2)},\ \mathbb{E}(\|\mathcal{R}\widetilde{f_{K,L_n}}-\mathcal{R}f_{K,L_n}\|^2)\le D_K/n,\ \ \mathbb{E}(\|\widehat{f_{K,L_n}}-\widetilde{f_{K,L_n}}\|^2)\le E_K/n.$$

Below, we prove using the Talagrand Inequality that

$$\mathbb{E}\left(\sup_{t\in S_{\ell\vee\hat\ell}, \|t\|=1}\langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)}\rangle^2 - p(\ell,\hat\ell)\right)_+ \le \frac{C'}{n}, \tag{35}$$

where $p(\ell,\ell') = 8\ell\vee\ell'/n$ and $16p(\ell,\ell')\le \mathrm{pen}(\ell)+\mathrm{pen}(\ell')$ as soon as $\kappa\ge\kappa_0 = 16\times 8$.

Thus, we get $\mathbb{E}(16p(\ell,\hat\ell)-\mathrm{pen}(\hat\ell))\le \mathrm{pen}(\ell)$ and

$$\mathbb{E}(\|\widehat{f_{K,\hat\ell}}-f\|^2)\le 4\|f-f_\ell\|^2 + 4\mathrm{pen}(\ell) + 32A_K\Delta^{2(K+2)} + 32\frac{B_K}{n} + \frac{32C'}{n}.$$

Proof of (35). We consider $t\in S_{\ell^*}$ for $\ell^* = \ell\vee\ell'$ with $\ell,\ell'\le L_n$ and (see (26) and (27))

$$\nu_n(t) = c_1(\Delta)\langle t, \hat g_{L_n} - g_{L_n}\rangle = \frac{1}{n}\sum_{k=1}^n(\psi_t(Z_k) - \mathbb{E}(\psi_t(Z_k))$$

where

$$\psi_t(z) = \frac{c_1(\Delta)}{2\pi}\int t^*(u)e^{iuz}du = c_1(\Delta)t(z).$$

We apply the Talagrand Inequality recalled in Section 8, and to this aim, we compute the quantities M, H, v. First

$$\sup_{t\in S_{\ell^*},\|t\|=1}\sup_z|\psi_t(z)|\le \frac{c_1(\Delta)}{2\pi}\sqrt{2\pi\ell^*}\times \sup_{t\in S_{\ell^*},\|t\|=1}\|t^*\| = c_1(\Delta)\sqrt{\ell^*} := M.$$

The density of Z_1 is g which satisfies

$$\|g\|_\infty \le \sum_{m\ge 1}\frac{1}{e^{c\Delta}-1}\frac{(c\Delta)^m}{m!}\|f^{\star\, m}\|_\infty \le \|f\|_\infty.$$

Therefore,

$$\sup_{t\in S_{\ell^*},\|t\|=1}\mathrm{Var}(\psi_t(Z_1))\le c_1^2(\Delta)\times\sup_{t\in S_{\ell^*},\|t\|=1}\mathbb{E}(t^2(Z_1))\le c_1^2(\Delta)\|f\|_\infty := v.$$

Lastly, using the bound in (29) and the fact that for $t\in S_{\ell^*}$,

$$\langle t, \widetilde{f_{K,L_n}}^{(1)} - f_{K,L_n}^{(1)}\rangle = \langle t, \widetilde{f_{K,\ell^*}}^{(1)} - f_{K,\ell^*}^{(1)}\rangle,$$

we get

$$\begin{aligned}\mathbb{E}(\sup_{t\in S_{\ell^*},\|t\|=1}\nu_n^2(t)) &= \mathbb{E}\left(\sup_{t\in S_{\ell^*},\|t\|=1}\langle t,\widetilde{f_{K,\ell^*}}^{(1)}-f_{K,\ell^*}^{(1)}\rangle^2\right)\leq \mathbb{E}\left(\|\widetilde{f_{K,\ell^*}}^{(1)}-f_{K,\ell^*}^{(1)}\|^2\right)\\ &\leq \frac{2\ell^*}{n}:=H^2.\end{aligned}$$

Therefore, Lemma 8.1 yields with $\epsilon^2=1/2$,

$$\mathbb{E}(\sup_{t\in S_{\ell^*},\|t\|=1}\nu_n^2(t)-4H^2)\leq \frac{A_1}{n}(e^{-A_2\ell^*}+e^{-A_3\sqrt{n}})$$

for constants A_1,A_2,A_3 depending on $c_1(\Delta)$ and $\|f\|_\infty$. Now since $\sum_{\ell'=1}^{L_n}e^{-A_2\ell\vee\ell'}=\ell e^{-A_2\ell}+\sum_{\ell<\ell'\leq L_n}e^{-A_2\ell'}$ is bounded by say B_2 and $L_n e^{-A_3\sqrt{n}}$ is bounded by B_3, we get

$$\mathbb{E}\left(\sup_{t\in S_{\ell\vee\hat\ell},\|t\|=1}\nu_n^2(t)-8\frac{\ell\vee\hat\ell}{n}\right)\leq\sum_{\ell'}\mathbb{E}(\sup_{t\in S_{\ell\vee\ell'},\|t\|=1}\nu_n^2(t)-4H^2)\leq\frac{B_4}{n}.$$

This ends the proof of (35) and thus of Theorem 4.1. □

8. Appendix.

The result below follows from the Talagrand concentration inequality given in Klein and Rio (2005) and arguments in Birgé and Massart (1998) (see the proof of their Corollary 2 page 354).

Lemma 8.1. *(Talagrand Inequality) Let $Y_1,\dots,Y_n$ be independent random variables, let $\nu_{n,Y}(f)=(1/n)\sum_{i=1}^n[f(Y_i)-\mathbb{E}(f(Y_i))]$ and let $\mathcal{F}$ be a countable class of uniformly bounded measurable functions. Then for $\epsilon^2>0$*

$$\mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|^2-2(1+2\epsilon^2)H^2\Big]_+\leq\frac{4}{K_1}\left(\frac{v}{n}e^{-K_1\epsilon^2\frac{nH^2}{v}}+\frac{98M^2}{K_1n^2C^2(\epsilon^2)}e^{-\frac{2K_1C(\epsilon^2)\epsilon}{7\sqrt2}\frac{nH}{M}}\right),$$

with $C(\epsilon^2)=\sqrt{1+\epsilon^2}-1$, $K_1=1/6$, and

$$\sup_{f\in\mathcal{F}}\|f\|_\infty\leq M,\quad \mathbb{E}\Big[\sup_{f\in\mathcal{F}}|\nu_{n,Y}(f)|\Big]\leq H,\ \sup_{f\in\mathcal{F}}\frac1n\sum_{k=1}^n\mathrm{Var}(f(Y_k))\leq v.$$

By standard density arguments, this result can be extended to the case where $\mathcal{F}$ is a unit ball of a linear normed space, after checking that $f\mapsto\nu_n(f)$ is continuous and $\mathcal{F}$ contains a countable dense family.

References

[1] Basawa, I. V. ; Brockwell, P. J. (1982) Nonparametric estimation for nondecreasing Lévy processes. *J. Roy. Statist. Soc. Ser.* **B** 44, 262-269.

[2] Buchmann, B. and Grübel, R. (2003) Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.* **31** 1054-1074.

[3] Buchmann, B. (2009) Weighted empirical processes in the nonparametric inference for Lévy processes. *Math. Methods Statist.* **18**, 281-309.
[4] Chen, S. X., Delaigle, A. and Hall, P. (2010) Nonparametric estimation for a class of Lévy processes. *J. Econometrics* **157**, 257-271.
[5] Chesneau, C., Comte, F. and Navarro, F. (2013) Fast nonparametric estimation for convolutions of densities. Preprint MAP5.
[6] Comte, F. and Genon-Catalot (2009) Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications* **119**, 4088-4123.
[7] Comte, F. and Genon-Catalot, V. (2010) Nonparametric adaptive estimation for pure jump Lévy processes. *Ann. Inst. Henri Poincar Probab. Stat.* **46**, 595-617.
[8] Comte, F. and Genon-Catalot, V. (2011) Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.* **39**, 803-837.
[9] Duval, C. (2012a) Adaptive wavelet estimation of a compound Poisson process. Preprint arXiv:1203.3135.
[10] Duval, C. (2012b) Nonparametric estimation of a renewal reward process from discrete data. Preprint arXiv:1207.1611.
[11] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). Modelling extremal events. Berlin: Springer. MR1458613.
[12] Figueroa-López, J. E. and Houdré, C. (2006) Risk bounds for the non-parametric estimation of Lévy processes. *High dimensional probability*, 96-116, IMS Lecture Notes Monogr. Ser., 51, Inst. Math. Statist., Beachwood, OH.
[13] Figueroa-López, J. E. (2009) Nonparametric estimation of Lévy models based on discrete-sampling. Optimality, 117-146, IMS Lecture Notes Monogr. Ser., 57, Inst. Math. Statist., Beachwood, OH.
[14] Gugushvili, S. (2009) Nonparametric estimation of the characteristic triplet of a discretely observed Lévy process. *J. Nonparametr. Stat.* **21**, 321-343.
[15] Gugushvili, S. (2012) Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, 282-307.
[16] Jongbloed, G., van der Meulen, F. H. and van der Vaart, A. W. (2005) Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli* **11**, 759-791.
[17] Katz, R. W. (2002) Stochastic modeling of hurricane damage. *J. Appl. Meteorol.* **41**, 754-762.
[18] Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand J Stat.* **24**, 211-229.
[19] Kim, Y. (1999). Nonparametric Bayesian estimators for counting processes. *Ann. Statist.* **27**, 562-588.
[20] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.
[21] Neumann, Michael H. and Reiss, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15**, 223-248.
[22] Scalas, E. (2006). The application of continuous-time random walks in finance and economics. *Physica A* **362**, 225-239.
[23] Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, **16**, 925-935.
[24] Tsybakov, A. (2009) Introduction to Nonparametric Estimation. Springer.
[25] 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**, 967-989.
[26] Zhao, Z. and Wu, W. B. (2009). Nonparametric inference of discretely sampled stable Lévy processes. *J. Econometrics* **153**, 83-92.