HAL
archives-ouvertes.fr

Estimation of the Jump Size Density in a Mixed Compound Poisson Process

Fabienne Comte, Céline Duval, V Genon-Catalot, J Kappus

To cite this version:

Fabienne Comte, Céline Duval, V Genon-Catalot, J Kappus. Estimation of the Jump Size Density in a Mixed Compound Poisson Process. Scandinavian Journal of Statistics, Wiley, 2015, 42 (4), pp.1023-1044. <10.1111/sjos.12149>. <hal-01228389>

HAL Id: hal-01228389

https://hal.archives-ouvertes.fr/hal-01228389

Submitted on 13 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESTIMATION OF THE JUMP SIZE DENSITY IN A MIXED COMPOUND POISSON PROCESS.

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

Abstract. In this paper, we consider a mixed compound Poisson process, *i.e.* a random sum of *i.i.d.* random variables where the number of terms is a Poisson process with random intensity. We study nonparametric estimators of the jump density by specific deconvolution methods. First, assuming that the random intensity has exponential distribution with unknown expectation, we propose two types of estimators based on the observation of an *i.i.d.* sample. Risks bounds and adaptive procedures are provided. Then, with no assumption on the distribution of the random intensity, we propose two nonparametric estimators of the jump density based on the joint observation of the number of jumps and the random sum of jumps. Risks bounds are provided, leading to unusual rates for one of the two estimators. The methods are implemented and compared via simulations. February 25, 2015

Keywords. Adaptive methods, deconvolution, mixed compound Poisson process, nonparametric density estimation, penalization method.

1. Introduction

Compound Poisson processes are commonly used in many applied fields, especially in queuing and risk theory (see *e.g.* Embrechts *et al.* (1997), Grandell (1997), Mikosch (2009)). The model can be described as follows. Consider a Poisson process $(N(t))$ with intensity 1, $(\xi_i, i \geq 1)$ a sequence of *i.i.d.* random variables with common density f independent of N and λ a positive number. Then, $(N(\lambda t), t \geq 0)$ is a Poisson process with intensity λ and $X^\lambda(t) = \sum_{i=1}^{N(\lambda t)} \xi_i$ is a compound Poisson process with jump size density f. The process X^λ has independent and stationary increments and is therefore a special case of Lévy process with Lévy density λf. Lots of references on Lévy density estimation are available (see Comte and Genon-Catalot (2009), Figueroa-Lopez (2009), Neumann and Reiss (2009), Ueltzhöfer and Klüppelberg (2011), Gugushvili (2012)). Inference is generally based on a discrete observation of one sample path with sampling interval Δ and uses the n-sample of *i.i.d.* increments $(X^\lambda(k\Delta) - X^\lambda((k-1)\Delta), k \leq n)$. For the special case of compound Poisson process, van Es *et al.* (2007) build a kernel type estimator of f in the low frequency setting (Δ fixed), assuming that the intensity λ is known. In

[1]MAP5, UMR CNRS 8145, Université Paris Descartes.

[2]Institut für Mathematik, Rostock Universität.

Duval (2013) and in Comte *et al.* (2014), the same problem is considered with λ unknown and in the high frequency setting ($\Delta = \Delta_n$ tends to 0 while $n\Delta$ tends to infinity).

In this paper, we consider the case where the intensity λ is not deterministic but random. The model is now as follows. Let Λ be a positive random variable, independent of $(N(t), t \geq 0)$ and of the sequence $(\xi_i, i \geq 1)$. Then,

$$Y(t) = \sum_{i=1}^{N(\Lambda t)} \xi_i \tag{1}$$

defines a mixed compound Poisson process (see Grandell (1997)). Given that $\Lambda = \lambda$, the conditional distribution of $(Y(t))$ is identical to the distribution of $(X^\lambda(t))$. The mixed compound Poisson model belongs to the more general class of mixed effects models where some parameters are (unobserved) random variables. Mixed effects models which are extremely popular, are often used in studies in which repeated measurements are taken on a series of individuals (see *e.g.* Davidian and Giltinan (1995), Pinhero and Bates (2000), Antonio and Beirlant (2007), Belomestny (2011)). The randomness of parameters allows to account for the variability existing between subjects.

Model (1) belongs to the general class of random sums of random variables, widely used in a huge number of applications. For instance, in the field of non life insurance mathematics, companies hold portfolios of n contracts or policies. For each contract j, they know, per year ($\Delta = 1$), the total claim amount $\sum_{i=1}^{M_j} \xi_i^j$ where M_j is the number of claims (e.g. accidents) and ξ_i^j is the claim size of the $i-th$ claim of the $j-th$ contract. Independence of contracts is generally assumed. Modeling M_j as a mixed Poisson variable is also of common use (see Mikosch(2009) and Grandell (1997)). Note that an insurance company may observe each individual claim and the time the claim happens. Thus claim intensity and claim sizes can be estimated separately by standard statistics. Nevertheless, it is interesting to compare estimators based on highly detailed data (with exact claim amounts and times) to the ones proposed here which correspond to aggregation over time or end-of-year data alone.

The model of random sums is also used in stochastic activity of neurons for instance to model the amplitude of the endplate potential (see Tuckwell (1998), vol.2 chap.9). In such data, only aggregated data are available, so that our methods provide an adequate solution in this context.

Note that, in Boegsted and Pitts (2010), the authors assume that the distribution of M_j is general but known and construct a nonparametric plug in estimator of f. Below, we assume that the distribution of Λ either contains unknown parameters or is completely unknown, so that the distribution $N(\Lambda\Delta)$ is unknown.

In this paper, we consider, for a given time Δ, *i.i.d.* observations $(Y_j(\Delta), j = 1, \ldots, n)$ of $Y(\Delta)$ and our aim is to define and study nonparametric estimators of f. Note that, for deterministic λ, the n-sample of increments $(X^\lambda(k\Delta) - X^\lambda((k-1)\Delta), k \leq n)$ for one trajectory has exactly

the same distribution as an *i.i.d.* sample $(X_j^\lambda(\Delta), j = 1, \dots, n)$ for n trajectories at one instant Δ. Hence, the performances of estimation procedures based on *i.i.d.* data $(Y_j(\Delta), j = 1, \dots, n)$ may be compared with those of procedures based on increments $(X^\lambda(k\Delta) - X^\lambda((k-1)\Delta), k \le n)$ for one trajectory.

To fix notations, let $(\Lambda_j, j \ge 1)$ be *i.i.d.* with distribution $\nu(d\lambda)$, let $(N_j(t), j \ge 1)$ be *i.i.d.* Poisson processes with intensity 1 independent of $(\Lambda_j, j \ge 1)$ and consider, for $\Delta > 0$, the n-sample $(Y_j(\Delta) = \sum_{i=1}^{N_j(\Lambda_j\Delta)} \xi_i^j, j = 1, \dots, n)$ where $(\xi_i^j, j, i \ge 1)$ are *i.i.d.* with density f, and the sequence $(\xi_i^j, j, i \ge 1)$ is independent of $(\Lambda_j, (N_j(t)), j \ge 1)$. The paper is divided in two distinct parts, one is semi-parametric (Section 2) and the other purely nonparametric (Section 3). In both parts, our approach relies on deconvolution and requires the assumption that f belongs to $\mathbb{L}^2(\mathbb{R})$.

In Section 2, we assume that the observation is $(Y_j(\Delta), j \le n)$ and that the unobserved random intensities Λ_j's have an exponential distribution with unknown expectation μ. We define two different nonparametric estimators of f. First (method 1), introducing

$$Q_\Delta(u) := \mathbb{E}(e^{iuY(\Delta)}\mathbf{I}_{Y(\Delta)\neq 0}), \quad \phi_\Delta(u) := \mathbb{E}(e^{iuY(\Delta)}) \text{ and } q_\Delta := \mathbb{P}(Y(\Delta) \neq 0), \tag{2}$$

we observe that the Fourier transform $f^*(u)$ of f satisfies $f^*(u) = Q_\Delta(u)/(q_\Delta\phi_\Delta(u))$. This relation is specific to the case of Λ having an exponential distribution. We deduce an estimator $\hat{f}^*(u)$ of $f^*(u)$ based on empirical estimators of $Q_\Delta(u), \phi_\Delta(u), q_\Delta$. Then, by Fourier inversion, we build a collection of nonparametric estimators $\hat{f}_m(x)$ of f associated with a cut off parameter m and propose a data driven selection $\hat{m}$ of m leading to an adaptive estimator (Proposition 2, Theorem 1). Non asymptotic risk bounds are proved.
However, if Δ gets too small, the previous method deteriorates as q_Δ becomes small and $1/q_\Delta$ is badly estimated (this is pointed out on simulations). This is why we investigate a second method (method 2) which performs well for small Δ. The idea is that for $\mu\Delta < 1$, the following series development holds:

$$f^*(u) = \sum_{k\ge 0}(1+\mu\Delta)(-\mu\Delta)^k(g_\Delta^*(u))^{k+1} = \sum_{k=0}^{K}(1+\mu\Delta)(-\mu\Delta)^k(g_\Delta^*(u))^{k+1} + O(\Delta^{K+1}), \tag{3}$$

where g_Δ is the conditional density of $Y(\Delta)$ given $Y(\Delta) \neq 0$. We apply a method comparable to the one developed in Comte *et al.* (2014) for non random intensity. Using estimators of g_Δ^* and μ, we obtain an estimator of $f^*(u)$ depending on a truncation parameter K. We build by Fourier inversion estimators $(\hat{f}_{m,K})_m$ and propose a data-driven choice $\hat{m}_K$ of the cut-off m (see Proposition 3 and Theorem 2).

In Section 3, we no longer assume that Λ has exponential distribution. For identifiability purpose, we enrich the observation and assume that, in addition to $(Y_j(\Delta))$, the sample $(N_j(\Lambda_j\Delta), j = 1, \dots, n)$ is observed. We do not assume that Λ admits a density and the method works for deterministic (unknown) Λ. Here also, we define two estimators of f. The first one

(method 3) is based on the following idea. Assuming that $f^*(u) \neq 0$ for all u, we check that

$$\psi(u) := \frac{(f^*(u))'}{f^*(u)} = i\frac{G_\Delta(u)}{H_\Delta(u)}, \quad f^*(u) = \exp\left(\int_0^u \psi(v)dv\right) \tag{4}$$

where

$$G_\Delta(u) = \mathbb{E}\left(\frac{Y(\Delta)}{\Delta}e^{iuY(\Delta)}\right), \quad H_\Delta(u) = \mathbb{E}\left(\frac{N(\Lambda\Delta)}{\Delta}e^{iuY(\Delta)}\right). \tag{5}$$

Therefore, $\psi(u)$ can be estimated by using empirical counterparts of $G_\Delta(u), H_\Delta(u)$ leading to an estimator of $f^*(u)$. Afterwards, we proceed by deconvolution to define a collection of estimators $\widetilde{f}_m$ depending on a cut off parameter m. Proposition 4 gives the bound of the $\mathbb{L}^2$-risk of $\widetilde{f}_m$ for fixed m. The risk bounds are non standard as well as the proof to obtain them and give rise to unusual rates on standard examples when making the bias-variance trade-off. We propose an heuristic penalization criterion to define a data-driven choice of the cut off parameter.

The second estimator of this part (method 4) is based on the fact that each time that $N_j(\Lambda_j\Delta) = 1$, a random variable ξ is observed, or equivalently, the conditional distribution of $Y_j(\Delta)$ given $N_j(\Lambda_j\Delta) = 1$ is exactly f.

Section 4 illustrates our methods on simulated data for different examples of jump densities f and of distributions for Λ. It appears clearly that method 1 of Section 2 performs well for all values of Δ except very small ones contrary to method 2, as expected from theoretical results. The methods of Section 3 (methods 3 and 4) perform well for all Δ and for various distributions for Λ, including the exponential. A complete discussion on numerical results is given. Section 5 contains the main proofs. Additional proofs and numerical results are given in Supplementary material.

2. Semi-parametric strategies of estimation

In this section, we assume that Λ has an exponential distribution with parameter μ^{-1}.

2.1. Parameter estimation when Λ is $\mathcal{E}(\mu^{-1})$.

For any distribution $\nu(d\lambda)$ of Λ_j, the distribution of $N_j(\Lambda_j\Delta)$ is given by:

$$\mathbb{P}(N_j(\Lambda_j\Delta) = m) = \int_0^{+\infty} e^{-\lambda\Delta}\frac{(\lambda\Delta)^m}{m!}\nu(d\lambda), m \geq 0. \tag{6}$$

When Λ_j has density $\mu^{-1}e^{-\lambda\mu^{-1}}\mathbb{1}_{\lambda>0}$, the computation is explicit:

$$\mathbb{P}(N_j(\Lambda_j\Delta) = m) \;=\; \left(\frac{\mu\Delta}{\mu\Delta+1}\right)^m \frac{1}{\mu\Delta+1} := \alpha_m(\mu,\Delta),\ m \geq 0. \tag{7}$$

Noting that

$$\mathbb{P}(Y_j(\Delta) \neq 0) = \mathbb{P}(N_j(\Lambda_j\Delta) \neq 0) = 1 - \alpha_0(\mu,\Delta) = 1 - \frac{1}{1+\mu\Delta} = \frac{\mu\Delta}{1+\mu\Delta} := q_\Delta, \tag{8}$$

we get $\mu = \Delta^{-1} q_\Delta/(1-q_\Delta)$. To estimate $1/q_\Delta$ and μ, we define for k a positive constant

$$\hat{q}_\Delta = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{Y_j(\Delta)\neq 0}, \quad \frac{1}{\tilde{q}_\Delta} = \frac{1}{\hat{q}_\Delta}\mathbb{I}_{\hat{q}_\Delta \geq k\sqrt{\Delta/n}}, \quad \tilde{\mu} = \frac{1}{\Delta}\frac{\hat{q}_\Delta}{1-\hat{q}_\Delta}\mathbb{I}_{1-\hat{q}_\Delta \geq k\sqrt{\Delta/n}}. \tag{9}$$

The following properties hold.

Proposition 1. *Under the Assumption:*

[**A**] The parameter μ belongs to a compact interval $[\mu_0, \mu_1]$ with $\mu_0 > 0$,

and if $n\Delta \geq 1$, the estimators $\hat{q}_\Delta$, $1/\tilde{q}_\Delta$ and $\tilde{\mu}$ given by (9) satisfy for all integer $p \geq 1$,

$$\mathbb{E}(\hat{q}_\Delta - q_\Delta)^{2p} \leq C(p,\mu_1)\left(\frac{\Delta}{n}\right)^p, \quad \mathbb{E}\left(\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\right)^{2p} \leq C'(p,\Delta)\left(\frac{1}{n\Delta^3}\right)^p,$$

and

$$\mathbb{E}(\tilde{\mu} - \mu)^{2p} \leq \frac{C''(p,\Delta)}{(n\Delta)^p} \tag{10}$$

where $C(p,\mu_1)$ only depends on p and μ_1, $C'(p,\Delta) = C'(p) + O(\Delta)$, $C''(p,\Delta) = C''(p) + O(\Delta)$, and $C'(p,\Delta)$, $C''(p,\Delta)$ only depend on p, μ_0, μ_1 and Δ.

Assumption [**A**] is needed but definition (9) does not require the knowledge of μ_0, μ_1.

2.2. **Notation.** The following notations are used below. For $u : \mathbb{R} \to \mathbb{C}$ integrable, we denote its $\mathbb{L}^1$ norm and its Fourier transform respectively by

$$\|u\|_1 = \int_{\mathbb{R}} |u(x)|dx, \quad u^*(y) = \int_{\mathbb{R}} e^{iyx}u(x)dx, \quad y \in \mathbb{R}.$$

When u, v are square integrable, we denote the $\mathbb{L}^2$ norm and the $\mathbb{L}^2$ scalar product by

$$\|u\| = \left(\int_{\mathbb{R}} |u(x)|^2 dx\right)^{1/2}, \quad \langle u, v\rangle = \int_{\mathbb{R}} u(x)\overline{v}(x)dx \quad \text{with } z\overline{z} = |z|^2.$$

We recall that, for any integrable and square-integrable functions u, u_1, u_2, the following relations hold: $(u^*)^*(x) = 2\pi u(-x)$ and $\langle u_1, u_2\rangle = (2\pi)^{-1}\langle u_1^*, u_2^*\rangle$. The convolution product of u, v is denoted by: $u \star v(x) = \int_{\mathbb{R}} u(y)\bar{v}(x-y)dy$.

2.3. **Estimation of f for fixed sampling interval (method 1).** In this section, we propose an estimator of f assuming that Δ is fixed (heuristically, $\Delta = 1$). For Q_Δ, ϕ_Δ defined in (2), we have

$$\phi_\Delta(u) = \int_0^{+\infty} \mu^{-1}e^{-\lambda\mu^{-1}} \exp\left(-\lambda\Delta(1-f^*(u))\right)d\lambda = \frac{1}{1+\mu\Delta(1-f^*(u))}.$$

Simple computations yield:

$$Q_\Delta(u) = \phi_\Delta(u) - \frac{1}{1+\mu\Delta} = \frac{\mu\Delta f^*(u)}{(1+\mu\Delta)(1+\mu\Delta(1-f^*(u)))}.$$

Solving for $f^*(u)$ yields the following formula

$$f^*(u) = \frac{1+\mu\Delta}{\mu\Delta}\frac{Q_\Delta(u)}{Q_\Delta(u)+\frac{1}{1+\mu\Delta}} = \frac{Q_\Delta(u)}{q_\Delta\phi_\Delta(u)}. \tag{11}$$

This formula which is very specific to the case of Λ_j having exponential distribution with expectation μ suggests to estimate $f^*(u)$ as follows:

$$\hat{f}^*(u) = \frac{\hat{Q}_\Delta(u)}{\tilde{q}_\Delta\widetilde{\phi}_\Delta(u)} \quad \text{with} \quad \hat{Q}_\Delta(u) = \frac{1}{n}\sum_{j=1}^n e^{iuY_j(\Delta)}\mathbb{I}_{Y_j(\Delta)\neq 0}, \tag{12}$$

$1/\tilde{q}_\Delta$ is defined by (9), and, for k a constant,

$$\frac{1}{\widetilde{\phi_\Delta}(u)} = \frac{1}{\widehat{\phi_\Delta}(u)}\mathbb{I}_{|\widehat{\phi_\Delta}(u)|\geq k/\sqrt{n}}, \quad \widehat{\phi_\Delta}(u) = \frac{1}{n}\sum_{j=1}^n e^{iuY_j(\Delta)}. \tag{13}$$

Then, we apply Fourier inversion to (12), but as $\hat{f}^*$ is not integrable, a cut off is required. We propose thus

$$\hat{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\hat{f}^*(u)du. \tag{14}$$

Then we can bound the mean-square risk of the estimator as follows.

Proposition 2. *Assume that Λ is $\mathcal{E}(\mu^{-1})$, $f \in \mathbb{L}^2(\mathbb{R})$ and that* **[A]** *holds. Then the estimator $\hat{f}_m$, for $m \leq n\Delta$, given by (14) and (12) satisfies*

$$\mathbb{E}(\|\hat{f}_m - f\|^2) \leq \|f - f_m\|^2 + \frac{1}{\pi n q_\Delta}\int_{-\pi m}^{\pi m}\frac{du}{|\phi_\Delta(u)|^2} + \frac{c}{n\Delta}$$

where f_m is such that $f_m^* = f^*\mathbb{I}_{[-\pi m,\pi m]}$ and where the constant c does not depend on n or Δ.

The bias term $\|f - f_m\|^2$ is decreasing with m while the variance term, of order m/n, is increasing with m; this illustrates that a standard bias-variance compromise has to be performed. If f belongs to the Sobolev ball defined by $\mathcal{S}(\alpha, L) = \{f \in \mathbb{L}^2(\mathbb{R}),\ \int |f^*(u)|^2(1+u^2)^\alpha du \leq L\}$,

$$\|f - f_m\|^2 = \frac{1}{2\pi}\int_{|u|\geq\pi m}|f^*(u)|^2du \leq \frac{L}{2\pi}(1+(\pi m)^2)^{-\alpha} \leq c_L m^{-2\alpha}.$$

Therefore, we find that, for $m = m_{opt} \asymp (n\Delta)^{1/(2\alpha+1)}$, $\mathbb{E}(\|\hat{f}_{m_{opt}} - f\|^2) = O((n\Delta)^{-2\alpha/(2\alpha+1)})$, which is a standard nonparametric rate.

We propose a data driven way of selecting m, and we proceed classically by mimicking the bias-variance compromise. Setting $\mathcal{M}_n = \{1, \ldots, n\Delta\}$, we select

$$\hat{m} = \arg\min_{m\in\mathcal{M}_n}\left(-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\right) \quad \text{where} \quad \widehat{\text{pen}}(m) = \kappa\frac{1}{\tilde{q}_\Delta}\frac{1}{2\pi n}\int_{-\pi m}^{\pi m}\frac{du}{|\widetilde{\phi_\Delta}(u)|^2}. \tag{15}$$

Then we can prove the following result.

Theorem 1. *Assume that Λ is $\mathcal{E}(\mu^{-1})$, $f \in \mathbb{L}^2(\mathbb{R})$ and that assumption* **[A]** *holds. Then for any numerical constant κ such that $\kappa \geq \kappa_0 = 96$, we have*

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \leq c \inf_{m \in \mathcal{M}_n} \left(\|f - f_m\|^2 + \frac{\kappa}{2\pi n q_\Delta} \int_{-\pi m}^{\pi m} \frac{du}{|\phi_\Delta(u)|^2} \right) + \frac{c'_\Delta}{n}$$

where c is a numerical constant ($c = 4$ would suit) and c'_Δ depends on Δ, μ_0, μ_1 and $\|f\|$.

The bounds of Proposition 2 and Theorem 1 are non asymptotic and hold for all n and Δ. However, if Δ gets too small, the method deteriorates because q_Δ becomes small and $1/q_\Delta$ is badly estimated. Moreover $c'_\Delta \propto \Delta^{-4}$.

Remark 1. Theorem 1 states that the estimator $\hat{f}_{\hat{m}}$ is adaptive as the bias-variance compromise is automatically realized. It also states that there is a minimal value κ_0 such that for all $\kappa \geq \kappa_0$, the adaptive risk bound holds. From our proof, we find $\kappa_0 = 96$, which is not optimal. Indeed, in simple models, a minimal value for κ_0 may be computed. For instance, Birgé and Massart (2007) prove that for Gaussian regression or white noise models, the method works for all $\kappa \geq \kappa_0 = 1 + \eta$, $\eta > 0$, and explodes for $\kappa = 1 - \eta$. Obtaining the minimal value in another context is not obvious. This is why it is customary, when using a penalized method, to calibrate the value κ in the penalty by preliminary simulations.

2.4. **Estimation of f for small sampling interval (method 2).** Now, we assume that $\Delta = \Delta_n$ tends to 0 and that $n\Delta$ tends to infinity. We use an approach for small sampling interval which is different from the previous one. We consider representation (3) for f^*, which holds for $\mu\Delta < 1$. To estimate the terms of the series, we use Proposition 1 above, and (17) together with Proposition A.1 stated and proved in Supplementary material and inspired by Chesneau *et al.* (2013).

The distribution of $Y_j(\Delta)$ is given by:

$$\mathbb{P}_{Y(\Delta)}(dx) = \alpha_0(\mu, \Delta)\delta_0(dx) + \sum_{m \geq 1} \alpha_m(\mu, \Delta) f^{\star\, m}(x)dx,$$

where $\alpha_m(\mu, \Delta), m \geq 0$ is defined by (7) and $f^{\star\, m}$ denotes the m-th convolution power of f. The conditional distribution of $Y(\Delta)$ given $Y(\Delta) \neq 0$ has density and Fourier transform given by (see (2), (8))

$$g_\Delta(x) = \frac{1}{q_\Delta} \sum_{k \geq 1} \alpha_k(\mu, \Delta) f^{\star\, k}(x), \quad g^*_\Delta(u) = \frac{Q_\Delta(u)}{q_\Delta}.$$

Using (7), as $\mu\Delta|f^*(u)|/(1 + \mu\Delta) < 1$, the Fourier transform of g_Δ is given by:

$$g^*_\Delta(u) = \left(\frac{\mu\Delta}{1 + \mu\Delta}\right)^{-1} \sum_{k \geq 1} \frac{1}{1 + \mu\Delta} \left(\frac{\mu\Delta}{1 + \mu\Delta}\right)^k (f^*(u))^k = \frac{f^*(u)}{1 + \mu\Delta(1 - f^*(u))}$$

Thus $|g^*_\Delta(u)| \leq |f^*(u)|$ which implies that

$$\|g_\Delta\| \leq \|f\|. \tag{16}$$

Solving for $f^*(u)$ yields: $f^*(u) = (1+\mu\Delta)g^*_\Delta(u)/(1+\mu\Delta g^*_\Delta(u))$. Now, if $\mu\Delta < 1$, then the development (3) holds and we have to estimate μ and g^*_Δ. For μ we use the estimator $\tilde{\mu}$ given in (9). To estimate $g^*_\Delta(u)$, we set, see (9) and (12):

$$\widetilde{g^*_\Delta}(u) = \frac{\widehat{g^*_\Delta}(u)}{\max\left(1, |\widehat{g^*_\Delta}(u)|\right)} \quad \text{with} \quad \widehat{g^*_\Delta}(u) = \frac{\hat{Q}_\Delta(u)}{\tilde{q}_\Delta}. \tag{17}$$

To estimate f^*, we plug $\widetilde{g^*_\Delta}$ and $\tilde{\mu}$ in (3) and truncate the series up to order K:

$$\hat{f}^*_K(u) = \sum_{k=0}^{K} (-1)^k (1+\tilde{\mu}\Delta)(\tilde{\mu}\Delta)^k (\widetilde{g^*_\Delta}(u))^{k+1}.$$

A practical choice of K is given below based on Proposition 3. Then, we proceed with Fourier inversion with cut off, to define

$$\hat{f}_{m,K}(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} \hat{f}^*_K(u) e^{-iux} du.$$

We can prove the following result.

Proposition 3. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$, that* **[A]** *holds, that Λ is $\mathcal{E}(\mu^{-1})$, and that $2\mu_1\Delta < 1$ and $\Delta < 1$. Then, for any $m \le n\Delta$, we have*

$$\mathbb{E}\left(\|\hat{f}_{m,K} - f\|^2\right) \le \|f_m - f\|^2 + 12\frac{(1+\mu\Delta)^3}{\mu}\frac{m}{n\Delta} + A(\mu_1\Delta)^{2K+2} + \frac{E_K}{n\Delta},$$

where f_m is such that $f^*_m = f^* \mathbf{I}_{[-\pi m, \pi m]}$, $A = 4\|f\|^2(1+\mu_1\Delta)^2/(1-\mu_1\Delta)^2$ and E_K is a constant depending on K, μ_0, μ_1 and $\|f\|$.

If $f \in \mathcal{S}(\alpha, L)$, $\|f - f_m\|^2 \le c_L m^{-2\alpha}$ and choosing $m = m_{opt} \asymp (n\Delta)^{1/(2\alpha+1)}$ implies

$$\mathbb{E}(\|\hat{f}_{m_{opt},K} - f\|^2) \le c_1 (n\Delta)^{-2\alpha/(2\alpha+1)} + c_2 \Delta^{2K+2}.$$

In practice, we choose K such that $\Delta^{2K+2} \le 1/(n\Delta)$, so that the third term in the risk bound of Proposition 3 is negligible. That is

$$K \ge K_0 := \frac{1}{2}\left(\frac{\log(n)}{|\log(\Delta)|} - 3\right). \tag{18}$$

Now, we have to select m in a data driven way. To that aim, we propose

$$\hat{m}_K = \arg\min_{m \in \{1,\dots,[n\Delta]\}} \left(-\|\hat{f}_{m,K}\|^2 + \widetilde{\mathrm{pen}}(m)\right), \quad \widetilde{\mathrm{pen}}(m) = \kappa' \frac{(1+\tilde{\mu}\Delta)^2}{\tilde{q}_\Delta}\frac{m}{n}.$$

For the choice of κ' in the penalty, we refer to Remark 1. The following holds.

Theorem 2. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$, that Λ is $\mathcal{E}(\mu^{-1})$, that* **[A]** *holds and that $2\mu_1\Delta < 1$. Then there exists a numerical constant κ'_0 such that, for all $\kappa' \ge \kappa'_0$, we have*

$$\mathbb{E}\left(\|\hat{f}_{\hat{m}_K,K} - f\|^2\right) \le c \inf_{m \in \{1,\dots,[n\Delta]\}} \left(\|f_m - f\|^2 + \kappa' \frac{(1+\mu\Delta)^3}{\mu}\frac{m}{n\Delta}\right) + A(\mu_1\Delta)^{2K+2} + \frac{E'_K}{n\Delta},$$

where c is a numerical constant, A is defined in Proposition 3 and E'_K is a constant depending on K, μ_0, μ_1 and $\|f\|$.

2.5. **The case where Λ has Gamma distribution.** The distribution (6) is also explicit when Λ has Gamma distribution, with parameters (a, μ^{-1}) together with Q_Δ, ϕ_Δ as defined by (2). Indeed, plugging the Gamma density in formula (6), we get

$$\alpha_m(a, \mu, \Delta) = \frac{\Gamma(a+m)}{\Gamma(a)m!}\left(\frac{1}{1+\mu\Delta}\right)^a\left(\frac{\mu\Delta}{1+\mu\Delta}\right)^m.$$

If a is integer, this is a negative binomial distribution $\text{bin}^-(a, 1/1+\mu\Delta)$. To estimate the distribution of Λ, only one relation is available, which is $\mathbb{P}(Y(\Delta) = 0) = (1+\mu\Delta)^{-a}$. Therefore, we cannot estimate the two parameters a and μ. For instance if a is known, methods 1 and 2 of Section 2 can be generalized, up to some additional but tedious computations.

3. Nonparametric strategies

In this section, we make no assumption on the distribution of Λ and turn to the estimation of f, using both samples $(Y_j(\Delta), N_j(\Lambda_j\Delta))_j$. Observe that we cannot estimate f on the basis of the sample $(Y_j(\Delta))_j$ as in Section 2. Even in the case where the distribution of Λ is known up to two parameters (see Section 2.5), there occurs an identifiability problem.

3.1. **Method 3.** We start from the characteristic function and for ν denoting the distribution of Λ, we have, from (2),

$$\phi_\Delta(u) = \int_0^{+\infty} \exp\left(-\lambda\Delta(1 - f^*(u))\right)\nu(d\lambda)$$

and by derivation (see (5)),

$$iG_\Delta(u) = (f^*(u))'K_\Delta(u), \quad \text{where} \quad K_\Delta(u) = \mathbb{E}(\Lambda e^{iuY(\Delta)}). \tag{19}$$

For fixed $\Lambda = \lambda$, we get

$$\mathbb{E}\left(\frac{N(\lambda\Delta)}{\Delta}e^{iu\sum_{k=1}^{N(\lambda\Delta)}\xi_k}\right) = f^*(u)\mathbb{E}\left(\lambda e^{iu\sum_{k=1}^{N(\lambda\Delta)}\xi_k}\right)$$

and therefore

$$H_\Delta(u) = \mathbb{E}\left(\frac{N(\Lambda\Delta)}{\Delta}e^{iuY(\Delta)}\right) = f^*(u)K_\Delta(u).$$

We deduce that, if $H_\Delta \neq 0$, (4) holds. With the condition $f^*(0) = 1$, we obtain the formula

$$f^*(u) = \exp\left(\int_0^u \psi(v)dv\right),$$

where $\psi = (f^*)'/f^*$, see formula (4) (for $u \leq 0$, $\int_0^u = -\int_u^0$). We deduce an estimator by setting

$$\widetilde{f}_m(x) = \frac{1}{2\pi}\int_{-\pi m}^{\pi m} e^{-iux}\widetilde{f^*}(u)du \tag{20}$$

where

$$\widetilde{f^*}(u) = \widehat{f^*}(u)\mathbb{1}_{\{|\widehat{f^*}(u)|\leq 1\}} + \frac{\widehat{f^*}(u)}{|\widehat{f^*}(u)|}\mathbb{1}_{\{|\widehat{f^*}(u)|>1\}} = \frac{\widehat{f^*}(u)}{\max(1, |\widehat{f^*}(u)|)},$$

with, for some constant k,

$$\widehat{f^*}(u) = \exp\left(\int_0^u \widetilde{\psi}(v)dv\right), \quad \widetilde{\psi}(v) = i\frac{\hat{G}_\Delta(v)}{\tilde{H}_\Delta(v)}, \quad \hat{G}_\Delta(v) = \frac{1}{n\Delta}\sum_{j=1}^n Y_j(\Delta)e^{ivY_j(\Delta)},$$

$$\hat{H}_\Delta(v) = \frac{1}{n\Delta}\sum_{j=1}^n N_j(\Lambda_j\Delta)e^{ivY_j(\Delta)}, \quad \frac{1}{\tilde{H}_\Delta(v)} = \frac{1}{\hat{H}_\Delta(v)}\mathbb{1}_{\{|\hat{H}_\Delta(v)|\geq k(n\Delta)^{-1/2}\}}.$$

We introduce the following assumption depending on an integer p.

[**B**] (i) $\forall u \in \mathbb{R},\ f^*(u) \neq 0$, and there exists $K_0 > 0$ such that $\forall u \in \mathbb{R}$, $|K_\Delta(u)| \geq K_0$ (see (19)).

(ii)(p) $\mathbb{E}(\xi^{2p}) < +\infty$, $\mathbb{E}(\Lambda^{2p}) < +\infty$.

(iii) $\|G'_\Delta\|_1 < +\infty$.

To justify assumption [**B**](i), consider the case where Λ follows an exponential distribution $\mathcal{E}(\mu^{-1})$. Then

$$K_\Delta(u) = \frac{\mu}{[1+\mu\Delta(1-f^*(u))]^2} \quad \text{and} \quad K_\Delta(u) \sim_{u\to+\infty} \frac{\mu}{[1+\mu\Delta]^2}.$$

Thus H_Δ is not lower bounded near infinity contrary to $K_\Delta(u)$.

Under [**B**] (ii), $\mathbb{E}[(Y(\Delta))^{2p}] = \Delta\mathbb{E}(\Lambda)\mathbb{E}(\xi^{2p}) + o(\Delta)$. Indeed, we first compute the cumulants of the conditional distribution of $Y(\Delta)$ given Λ. Then we deduce the conditional moments using the link between moments and cumulants. Integrating with respect to Λ gives the result. Note that [**B**](ii) implies that G'_Δ exists, with

$$iG'_\Delta(u) = (f^*)''(u)K_\Delta(u) + i(f^*)'(u)\mathbb{E}(\Lambda Y(\Delta)e^{iuY(\Delta)}).$$

If $(f^*)'$ and $(f^*)''$ are integrable, [**B**](iii) holds. We can prove the following result.

Proposition 4. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$ and* [**B**] *hold. Let $\widetilde{f}_m$ be given by (20) and let Δ be fixed, $n\Delta \geq 1$. Then the following bound holds:*

$$\begin{aligned}\mathbb{E}(\|\widetilde{f}_m - f\|^2) \leq\ & \|f - f_m\|^2 + \frac{c_1}{n\Delta}\int_{-\pi m}^{\pi m}|f^*(u)|^2\left(\int_0^{|u|}\frac{dv}{|f^*(v)|^2} + \left(\int_0^{|u|}\frac{|(f^*)'(v)|}{|f^*(v)|^2}dv\right)^2\right)du \\ & + \frac{c_2}{(n\Delta)^p}\int_{-\pi m}^{\pi m}\left(1 + \left(\int_0^{|u|}\left|\frac{(f^*)'(v)}{f^*(v)}\right|^2 dv\right)^p\right)\left(\int_0^{|u|}\frac{dv}{|f^*(v)|^2}\right)^p du \\ & + \frac{c_3}{(n\Delta)^{2p-1}}\int_{-\pi m}^{\pi m}\left(\int_0^{|u|}\frac{dv}{|f^*(v)|}\right)^{2p} du \end{aligned} \tag{21}$$

where the constants c_i, $i = 1, 2, 3$ depend on $\|G'_\Delta\|_1$, K_0 and the moments of Λ and ξ_i up to order $2p$.

Note that the unknown function appears in the variance terms which is unusual in deconvolution problems.

3.2. **Rate of the estimator.** We study the resulting rate on different examples.

• Gamma distribution. Let $f \sim \Gamma(\alpha, 1)$. Then $f^*(u) = (1 - iu)^{-\alpha}$ and

$$(f^*)'(u)/f^*(u) = -\frac{i\alpha}{1 - iu}.$$

Note that Assumption **[B]**(iii) is fulfilled. We have $\|f - f_m\|^2 = O(m^{-2\alpha+1})$ so that $\alpha > 1/2$ is required for consistency of the estimator. For the variance terms, using the bound (21), we have

$$\mathbb{V}_1 := \int_{-\pi m}^{\pi m} |f^*(u)|^2 \left(\int_0^{|u|} \frac{dv}{|f^*(v)|^2} + \left(\int_0^{|u|} \frac{|(f^*)'(v)|}{|f^*(v)|^2} dv \right)^2 \right) du = O(m^2),$$

$$\mathbb{V}_2 := \int_{-\pi m}^{\pi m} \left(1 + \left(\int_0^{|u|} \left| \frac{(f^*)'(v)}{f^*(v)} \right|^2 dv \right)^p \right) \left(\int_0^{|u|} \frac{dv}{|f^*(v)|^2} \right)^p du = O(m^{(2\alpha+1)p+1}),$$

and

$$\mathbb{V}_3 := \int_{-\pi m}^{\pi m} \left(\int_0^{|u|} \frac{dv}{|f^*(v)|} \right)^{2p} du = O(m^{(2\alpha+1)p+1}).$$

Optimizing the bias and $\mathbb{V}_1/(n\Delta)$ yields $m_{opt,1} \asymp (n\Delta)^{1/(2\alpha+1)}$ and a rate $O((n\Delta)^{-(2\alpha-1)/(2\alpha+1)})$. Optimizing the bias and $\mathbb{V}_2/(n\Delta)^p$ yields $m_{opt,2} \asymp (n\Delta)^{1/(1+2\alpha(1+1/p))}$ and a rate

$$O((n\Delta)^{-(2\alpha-1)/(2\alpha+1+2\alpha/p)}).$$

Optimizing the bias and $\mathbb{V}_3/(n\Delta)^{2p-1}$ yields $m_{opt,3} \asymp (n\Delta)^{(2p-1)/(2p(\alpha+1)+2\alpha)}$ and a rate

$$O((n\Delta)^{-(2\alpha-1)(2p-1)/(2p(\alpha+1)+2\alpha)}).$$

For $p \geq 2$ the rate is of order $(n\Delta)^{-(2\alpha-1)/(2\alpha+1+\frac{2\alpha}{p})}$, which is close to $(n\Delta)^{-(2\alpha-1)/(2\alpha+1)}$ for large p. Thus, as p can be as large as desired, $\mathbb{V}_1$ and $\mathbb{V}_2$ get comparable.

• Gaussian distribution. Let us consider $f^*(u) = e^{-u^2/2}$, $(f^*)'(u)/f^*(u) = -u$. Assumption **[B]**(iii) is fulfilled. We use Lemma 4 recalled in Section 5.4, to derive $\|f - f_m\|^2 = O(m^{-1}e^{-(\pi m)^2})$, $\mathbb{V}_1 = O(m)$, $\mathbb{V}_2 = O(m^{2p-1}e^{p(\pi m)^2})$ and $\mathbb{V}_3 = O(m^{-2p-1}e^{p(\pi m)^2})$. We choose

$$(\pi m_{opt})^2 = \frac{p}{p+1} \log(n\Delta) - \frac{2p}{p+1} \log(\log(n\Delta)),$$

and get the rate

$$(n\Delta)^{-\frac{p}{p+1}} (\log n\Delta)^{\frac{p-1}{p+1}}.$$

Note that optimizing the bias and $\mathbb{V}_1/(n\Delta)$ leads to the rate $\sqrt{\log(n\Delta)}/(n\Delta)$. Here again, for large p, the two terms $\mathbb{V}_1$ and $\mathbb{V}_2$ are comparable.

3.3. **Cut off selection.** As previously, we need to propose a data-driven selection of the cut off. The bias is estimated up to a constant by $-\|\widetilde{f}_m\|^2$. The penalty is built by estimating the variance term of the risk bound. Here, we have three variance terms and we do not know which term to choose to build a penalty. Thus we select m as in method 1 (Section 2.3):

$$\widetilde{m} = \arg\min_m(-\|\widetilde{f}_m\|^2 + \widehat{\mathrm{pen}}(m))$$

where $\widehat{\mathrm{pen}}(m)$ is defined in (15) and built using the $Y_j(\Delta)$'s only. The numerical results confirm that this strategy works well.

3.4. **Method 4.** A simple procedure is available for estimating f based on the joint observation $(N_j(\Lambda_j\Delta), Y_j(\Delta))_{1\le j\le n}$. Note that

$$\mathbb{P}(N_j(\Lambda_j\Delta) = 1) := \alpha_1(\nu, \Delta) = \Delta \int_0^{+\infty} e^{-\lambda\Delta}\lambda\nu(d\lambda) > 0$$

and that the conditional distribution of $Y_j(\Delta)$ given $N_j(\Lambda_j\Delta) = 1$ is identical to the distribution of ξ_1^j. Hence, let us set, for k a positive constant:

$$\frac{1}{\tilde{\alpha}_1} = \frac{1}{\hat{\alpha}_1}\mathbb{I}_{\hat{\alpha}_1\ge k\sqrt{\Delta/n}}, \quad \hat{\alpha}_1 = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{(N_j(\Lambda_j\Delta)=1)}, \tag{22}$$

$$\check{f}_m(x) = \frac{1}{2\pi\tilde{\alpha}_1}\int_{-\pi m}^{\pi m} e^{-iux}\left(\frac{1}{n}\sum_{i=1}^n e^{iuY_j(\Delta)}\mathbb{I}_{(N_j(\Lambda_j\Delta)=1)}\right)du. \tag{23}$$

The following property holds.

Proposition 5. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$, $\mathbb{E}(\Lambda) < +\infty$, $\mathbb{E}(\Lambda e^{-\Lambda\Delta}) \ge k_0$ and $n\Delta \ge 1 \vee \frac{4k^2}{k_0^2}$. Then $\check{f}_m$ defined by (23) satisfies*

$$\mathbb{E}(\|\check{f}_m - f\|^2) \le \|f - f_m\|^2 + \frac{4m}{n\alpha_1(\nu,\Delta)}\left(1 + \frac{2k^2\Delta}{\alpha_1(\nu,\Delta)}\frac{1}{n\alpha_1(\nu,\Delta)}\right)$$

where f_m is such that $f_m^* = f^*\mathbb{I}_{[-\pi m,\pi m]}$

Note that $\alpha_1(\nu,\Delta) = \Delta(\mathbb{E}(\Lambda) + o(1))$. The variance term is of order $O(m/(n\Delta))$. We propose the following adaptive choice for m:

$$\check{m} = \arg\min_{m\le n\Delta}\left(-\|\check{f}_m\|^2 + \frac{\kappa''}{n\tilde{\alpha}_1}\right).$$

The proof of Proposition 5 follows the same lines as the analogous Proposition 2 and is omitted. The proof of adaptiveness of $\check{f}_{\check{m}}$ is also omitted.

The interest of this estimator is obviously its simplicity. However, it strongly depends on the observed value $n\hat{\alpha}_1$. If this value is too small, the estimator performs poorly, as illustrated in the numerical simulations.

4. Illustrations of the methods

In this section, we illustrate the estimators with data driven cut off on simulated data. The truncation constant k is always taken equal to 0.5 (equations (9) and (13)), except in (22) where $k = 0$. In all tables, the $\mathbb{L}_2$-risks are computed for f a Gaussian $\mathcal{N}(0, 3)$ and Λ either an exponential distribution with mean 1, or a uniform distribution on $[5, 6]$, a translated exponential $\mathcal{E}(2) + 5$, a translated $\mathcal{B}eta$ distribution $\mathcal{B}eta(2, 2) + 5$, a Gamma distribution $\Gamma(6, 3/2)$. To compute $\mathbb{L}_2$-risks, we perform 1000 Monte-Carlo iterations.

The cut off m is selected among 200 equispaced values between 0.01 and 2. All methods require the calibration of the constants $\kappa, \kappa', \kappa''$ in penalties. After preliminary experiments, we take $\kappa = 0.21$ (method 1) and $\kappa' = \kappa'' = 5$ (methods 2, 4).

Method 1 works for fixed values of Δ ($\Delta = 1, 2$), but also for small values (0.1 to 0.9). However, when Δ gets too small (0.01), the risk increases. On the other hand, method 2 completely fails for $\Delta = 1$, as predicted by the theory ($A = +\infty$ in the risk bound of Proposition 3 when $\mu\Delta = 1$); for $\Delta = 0.5, 0.9$, methods 1 and 2 have comparable risks while for $\Delta \leq 0.1$, method 2 is better. The cut off values are rather small and stable (standard deviations are of order 10^{-2}). The value K is taken of order $\sup(1, K_0)$ for K_0 defined in formula (18) (see Tables B.1, B.2 in Supplementary material).

In Figure 1, 50 estimated curves of a Gaussian mixture by methods 1 and 2 are plotted, for different sample sizes $n = 500, 2000, 5000$. The two methods distinguish well the two modes and are improved as n increases.

Table 1. Mean of the $\mathbb{L}_2$-risks for methods 1, 3 and 4. $\Delta = 1$, $\Lambda \sim \mathcal{E}(1)$ and f is $\mathcal{N}(0, 3)$; standard deviation in parenthesis.

n	Method 1	Method 3	Method 4
200	$5.6 \cdot 10^{-3}$ $(5.7 \cdot 10^{-3})$	$5.8 \cdot 10^{-3}$ $(5.9 \cdot 10^{-3})$	$6.0 \cdot 10^{-3}$ $(4.6 \cdot 10^{-3})$
500	$2.7 \cdot 10^{-3}$ $(2.5 \cdot 10^{-3})$	$2.9 \cdot 10^{-3}$ $(5.0 \cdot 10^{-3})$	$2.5 \cdot 10^{-3}$ $(2.2 \cdot 10^{-3})$
1000	$1.5 \cdot 10^{-3}$ $(1.7 \cdot 10^{-3})$	$1.8 \cdot 10^{-3}$ $(1.7 \cdot 10^{-3})$	$1.4 \cdot 10^{-3}$ $(1.1 \cdot 10^{-3})$
2000	$7.9 \cdot 10^{-4}$ $(5.7 \cdot 10^{-4})$	$1.1 \cdot 10^{-3}$ $(8.0 \cdot 10^{-4})$	$7.3 \cdot 10^{-4}$ $(5.5 \cdot 10^{-4})$
5000	$3.7 \cdot 10^{-4}$ $(2.9 \cdot 10^{-4})$	$5.3 \cdot 10^{-4}$ $(4.0 \cdot 10^{-4})$	$3.1 \cdot 10^{-4}$ $(2.3 \cdot 10^{-4})$

In Table 1 we compare method 1 to methods 3, 4 when Λ is exponential. They behave similarly, with good results even for small values of n. Method 4 performs surprisingly well and is stable. In Table 2, we change the distribution of Λ and therefore we show no results for method 1, since it does not work in that case, neither in theory nor in practice. The chosen

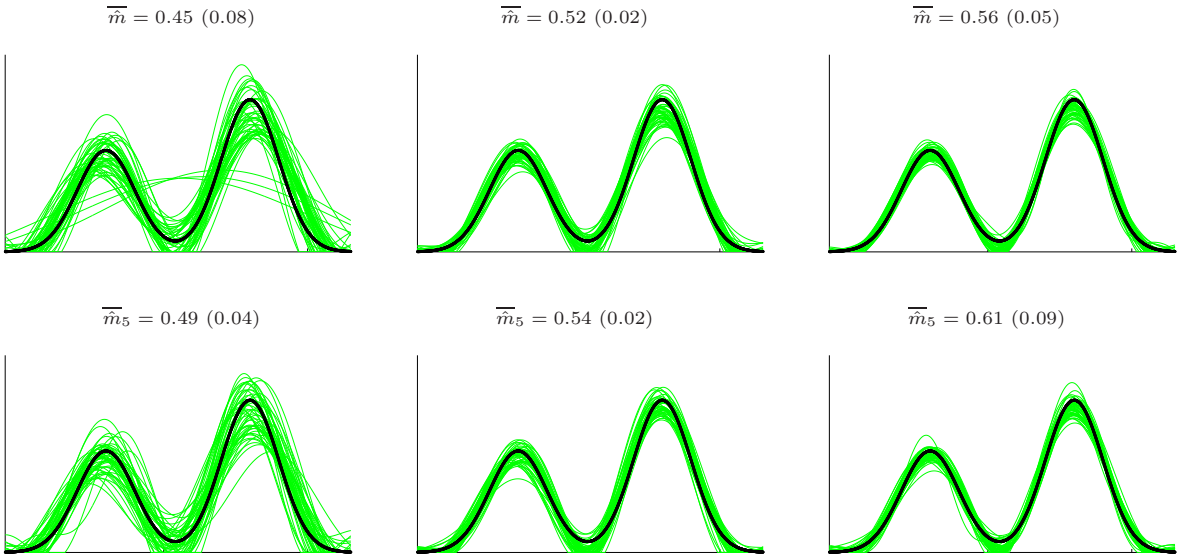

FIGURE 1. Estimation of f for a Gaussian mixture $0.4\mathcal{N}(-2,1)+0.6\mathcal{N}(3.1)$ for $n = 500$ (first column) $n = 2000$ (second column) and $n = 5000$ (third column) with method 1 (first line) and method 2 (second line, $K = 5$) for $\Delta = 1/2$. True density (bold black line) and 50 estimated curves (green lines). $\Lambda \sim \mathcal{E}(1)$. $\bar{\hat{m}}, \bar{\hat{m}}_5$ is the mean of cutoff values with s.d. in parenthesis.

distributions for Λ make method 4 perform worse than method 3. For $n \leq 1000$, method 3 better than method 4. For larger n, the methods become equivalent. Method 4 fails here because the number $n\hat{\alpha}_1$ (see (22)) is too small.

In Figure 2, 50 estimated curves of a Gumbel distribution by methods 1 and 3 are plotted, for different sample sizes $n = 500, 2000$, for Λ an exponential $\mathcal{E}(1)$ and a uniform $\mathcal{U}([1,2])$. Columns 1 and 2 allow to compare methods 1 and 3 when Λ is exponential. Method 3 has good performances without knowledge of the distribution of Λ. In all cases, the values of m are small and stable.

5. Main proofs

The proofs of Proposition 1, Proposition 3 and Theorem 2 are given in Supplementary material, as well as the proofs of Lemmas 1 and 2 stated hereafter.

First we state a useful lemma, similar to Lemma 2.1 p. 310 in Neumann (1997), proved in Section A.1. of Supplementary material.

Lemma 1. $\forall u \in \mathbb{R}, \quad \mathbb{E}(|\frac{1}{\widetilde{\phi_\Delta}(u)} - \frac{1}{\phi_\Delta(u)}|^{2p}) \leq c_p \left(\frac{1}{|\phi_\Delta(u)|^{2p}} \wedge \frac{n^{-p}}{|\phi_\Delta(u)|^{4p}} \right).$

TABLE 2. Mean of the $\mathbb{L}_2$-risks for methods 3 and 4; standard deviation in parenthesis. $\Delta = 1$, $\Lambda \sim \mathcal{U}([5,6])$, $\mathcal{E}(2)+5$, $\mathcal{B}eta(2,2)+5$, $\Gamma(6,3/2)$ and f is $\mathcal{N}(0,3)$.

Λ	n	Method 3	Method 4
$\mathcal{U}[5,6]$	200	$1.7 \cdot 10^{-2}$ $(1.3 \cdot 10^{-2})$	$5.6 \cdot 10^{-2}$ $(1.5 \cdot 10^{-2})$
	500	$1.1 \cdot 10^{-2}$ $(7.4 \cdot 10^{-3})$	$2.8 \cdot 10^{-2}$ $(1.7 \cdot 10^{-2})$
	1000	$8.7 \cdot 10^{-3}$ $(7.9 \cdot 10^{-3})$	$1.3 \cdot 10^{-2}$ $(9.7 \cdot 10^{-3})$
$\mathcal{B}eta(2,2)+5$	200	$1.6 \cdot 10^{-2}$ $(1.1 \cdot 10^{-2})$	$5.6 \cdot 10^{-2}$ $(1.6 \cdot 10^{-2})$
	500	$1.2 \cdot 10^{-2}$ $(8.5 \cdot 10^{-3})$	$2.9 \cdot 10^{-2}$ $(2.1 \cdot 10^{-2})$
	1000	$8.9 \cdot 10^{-3}$ $(7.4 \cdot 10^{-3})$	$1.3 \cdot 10^{-2}$ $(9.1 \cdot 10^{-3})$
$\mathcal{E}(2)+5$	200	$1.6 \cdot 10^{-2}$ $(1.2 \cdot 10^{-2})$	$5.6 \cdot 10^{-2}$ $(1.5 \cdot 10^{-2})$
	500	$1.1 \cdot 10^{-2}$ $(8.6 \cdot 10^{-3})$	$2.7 \cdot 10^{-2}$ $(1.6 \cdot 10^{-2})$
	1000	$8.3 \cdot 10^{-3}$ $(6.4 \cdot 10^{-3})$	$1.3 \cdot 10^{-2}$ $(8.5 \cdot 10^{-3})$
$\Gamma(6,3/2)$	200	$2.6 \cdot 10^{-2}$ $(1.4 \cdot 10^{-2})$	$6.3 \cdot 10^{-2}$ $(2.2 \cdot 10^{-2})$
	500	$1.8 \cdot 10^{-2}$ $(1.0 \cdot 10^{-2})$	$4.3 \cdot 10^{-2}$ $(1.8 \cdot 10^{-2})$
	1000	$1.5 \cdot 10^{-2}$ $(8.9 \cdot 10^{-3})$	$2.2 \cdot 10^{-2}$ $(1.3 \cdot 10^{-2})$

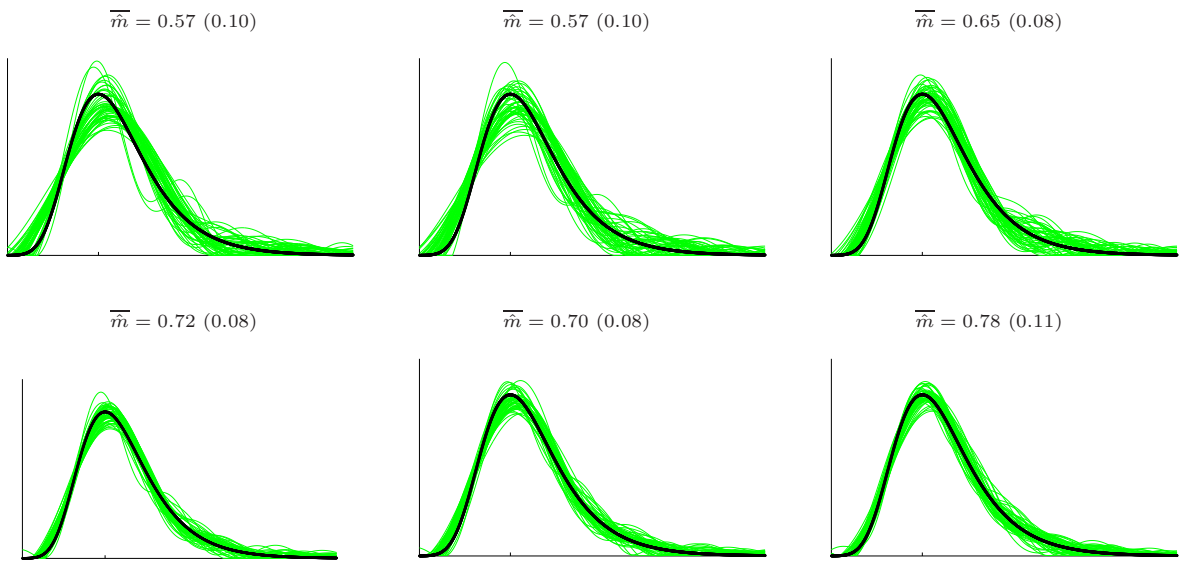

FIGURE 2. Estimation of f for a Gumbel distribution with c.d.f. $F(x) = \exp(-\exp(-x))$, $x > 0$. $n = 500$ (first line) $n = 2000$ (second line) with method 1 (first column) and method 3 (second and third column) for $\Delta = 1$. In the first two columns Λ is $\mathcal{E}(1)$ and in the third Λ is $\mathcal{U}[1,2]$. True density (bold black line) and 50 estimated curves (green lines). The value $\overline{\widehat{m}}$ is the mean over the 50 selected $\overline{\widehat{m}}$'s (with standard deviation in parenthesis).

5.1. **Proof of Proposition 2.** We use the fact that $|\phi_\Delta(u)|^{-1} \leq 1 + 2\mu\Delta$, $q_\Delta^{-1} \leq 1 + 1/(\mu_0\Delta)$, $\|f^*\|^2 = 2\pi\|f\|^2$, Proposition 1 and Lemma 1. Rosenthal's inequality (see Hand and Heyde (1980), p.23) implies for $p \geq 1$,

$$\mathbb{E}\left(|\hat{Q}_\Delta(u) - Q_\Delta(u)|^{2p}\right) \leq \frac{C(2p)}{n^{2p}}(nq_\Delta + (nq_\Delta)^p) \leq c_p(\mu_1 + \mu_1^p)\left(\frac{\Delta}{n}\right)^p. \tag{24}$$

Then we have the decomposition, using (11),

$$\hat{f}^*(u) - f^*(u) = T_0(u) + \sum_{i=1}^{6} T_i(u), \quad \text{with } T_0(u) = \frac{\hat{Q}_\Delta(u) - Q_\Delta(u)}{q_\Delta \phi_\Delta(u)}, \tag{25}$$

$$\begin{aligned}
T_1(u) =& \Big(\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\Big) q_\Delta f^*(u), \quad T_2(u) = \phi_\Delta(u) f^*(u)\Big(\frac{1}{\widetilde{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)}\Big),\\
T_3(u) =& Q_\Delta(u)\Big(\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\Big)\Big(\frac{1}{\widetilde{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)}\Big), \quad T_4(u) = \frac{\hat{Q}_\Delta(u) - Q_\Delta(u)}{\phi_\Delta(u)}\Big(\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\Big)\\
T_5(u) =& (\hat{Q}_\Delta(u) - Q_\Delta(u))\Big(\frac{1}{\widetilde{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)}\Big)\Big(\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\Big),\\
T_6(u) =& \frac{\hat{Q}_\Delta(u) - Q_\Delta(u)}{q_\Delta}\Big(\frac{1}{\widetilde{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)}\Big).
\end{aligned}$$

Then, $\displaystyle\int_{-\pi m}^{\pi m} |\hat{f}^*(u) - f^*(u)|^2 du \leq 2\int_{-\pi m}^{\pi m} |T_0(u)|^2 du + 12\sum_{i=1}^{6}\int_{-\pi m}^{\pi m} |T_i(u)|^2 du.$

First from (24), we have $\mathbb{E}\left(\int_{-\pi m}^{\pi m} |T_0(u)|^2 du\right) \leq \frac{1}{nq_\Delta}\int_{-\pi m}^{\pi m} \frac{du}{|\phi_\Delta(u)|^2}$. For the following bounds, we use constants c, c' which may change from line to line but depend neither on n nor on Δ. We have by Proposition 1

$$\mathbb{E}\left(\int_{-\pi m}^{\pi m} |T_1(u)|^2 du\right) = \int_{-\pi m}^{\pi m} |f^*(u)|^2 du \mathbb{E}\left(q_\Delta^2\Big|\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\Big|^2\right) \leq \frac{c}{n\Delta} 2\pi\|f\|^2,$$

and analogously by Lemma 1, $\mathbb{E}\left(\int_{-\pi m}^{\pi m} |T_2(u)|^2 du\right) \leq c'\|f\|^2/n$. Then as $m \leq n\Delta$, with both Lemma 1 and Proposition 1, we get

$$\mathbb{E}\left(\int_{-\pi m}^{\pi m} |T_3(u)|^2 du\right) \leq \frac{c}{n^2\Delta}\int_{-\pi m}^{\pi m} \frac{du}{|\phi_\Delta(u)|^4} \leq \frac{c'}{n}.$$

The three last terms are bounded in the same way. Remark that $\|\hat{f}_m - f\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2$ as $f - f_m$ and $\hat{f}_m - f_m$ have Fourier transforms with disjoint supports and that $\int_{-\pi m}^{\pi m} |\hat{f}^*(u) - f^*(u)|^2 du = 2\pi\|\hat{f}_m - f_m\|^2$. Gathering all the terms gives the result of Proposition 2.□

5.2. **Proof of Theorem 1.** Let $S_m = \{t \in \mathbb{L}^2(\mathbb{R}),\ t^* = t^* \mathbb{1}_{[-\pi m, \pi m]}\}$, and consider the contrast $\gamma_n(t) = \|t\|^2 - \frac{2}{2\pi}\langle t^*, \hat{f}^*\rangle$. Clearly, $\hat{f}_m = \arg\min_{t\in S_m} \gamma_n(t)$ and $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$. Moreover, we note that

$$\gamma_n(t) - \gamma_n(s) = \|t - f\|^2 - \|s - f\|^2 - \frac{2}{2\pi}\langle t^* - s^*, \hat{f}^* - f^*\rangle. \tag{26}$$

By definition of $\hat{m}$, we have $\gamma_n(\hat{f}_{\hat{m}}) + \widehat{\text{pen}}(\hat{m}) \leq \gamma_n(f_m) + \widehat{\text{pen}}(m)$. This with (26) implies

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq \|f - f_m\|^2 + \widehat{\text{pen}}(m) + \frac{2}{2\pi}\langle \hat{f}^*_{\hat{m}} - f^*_m, \hat{f}^* - f^*\rangle - \widehat{\text{pen}}(\hat{m}). \tag{27}$$

Writing that

$$\begin{aligned}
2\langle \hat{f}^*_{\hat{m}} - f^*_m, \hat{f}^* - f^*\rangle &\leq 2\|\hat{f}^*_{\hat{m}} - f^*_m\| \sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} |\langle t^*, \hat{f}^* - f^*\rangle| \\
&\leq \frac{1}{4}\|\hat{f}^*_{\hat{m}} - f^*_m\|^2 + 4 \sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} \langle t^*, \hat{f}^* - f^*\rangle^2 \\
&\leq \frac{1}{2}\|\hat{f}^*_{\hat{m}} - f^*\|^2 + \frac{1}{2}\|f^* - f^*_m\|^2 + 4 \sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} \langle t^*, \hat{f}^* - f^*\rangle^2,
\end{aligned}$$

plugging this in (27) and gathering the terms, we get

$$\frac{1}{2}\|\hat{f}_{\hat{m}} - f\|^2 \leq \frac{3}{2}\|f - f_m\|^2 + \widehat{\text{pen}}(m) + \frac{4}{2\pi} \sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} \langle t^*, \hat{f}^* - f^*\rangle^2 - \widehat{\text{pen}}(\hat{m}). \tag{28}$$

Now, we write the decomposition

$$\langle t^*, \hat{f}^* - f^*\rangle = \frac{1}{q_\Delta}\Big\langle t^*, \frac{\hat{Q}_\Delta - Q_\Delta}{\phi_\Delta}\Big\rangle + R(t)$$

where $R(t) = \sum_{i=1}^6 \langle t^*, T_i\rangle$ and the T_i's are defined by (25).

Clearly, the proof of Proposition 2, the Cauchy Schwarz inequality and $\|t^*\|^2 = 2\pi$ lead to $\mathbb{E}\left(\sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} \big|R(t)\big|^2\right) \leq c/(n\Delta)$. Thus, we have to study

$$\sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} q_\Delta^{-1}\big\langle t^*, (\hat{Q}_\Delta - Q_\Delta)/\phi_\Delta\big\rangle,$$

for which, we can prove (see the proof in A.3. of Supplementary material):

Lemma 2. *Under the Assumptions of Theorem 1, let*

$$p(m, \hat{m}) = \frac{3}{2\pi n q_\Delta} \int_{-\pi(m\vee\hat{m})}^{\pi(m\vee\hat{m})} \frac{du}{|\phi_\Delta(u)|^2}.$$

We have

$$\mathbb{E}\left(\frac{1}{2\pi} \sup_{t\in S_{\hat{m}\vee m}, \|t\|=1} \Big\langle t^*, \frac{\hat{Q}_\Delta - Q_\Delta}{q_\Delta \phi_\Delta}\Big\rangle^2 - p(m, \hat{m})\right)_+ \leq \frac{c}{n\Delta}.$$

Let us define

$$\Omega = \left\{ \left| \frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta} \right| \leq \frac{1}{2q_\Delta} \right\} \quad \text{and} \quad \text{pen}(m) = \frac{1}{2\pi n q_\Delta} \int_{-\pi m}^{\pi m} \frac{du}{|\phi_\Delta(u)|^2}.$$

We have $p(m,m') \leq 3\text{pen}(m) + 3\text{pen}(m')$ and on Ω, we have, $\forall m \in \mathcal{M}_n$,

$$\begin{aligned}
\mathbb{E}(\widehat{\text{pen}}(m)\mathbb{1}_\Omega) &\leq \frac{3}{2}\frac{\kappa}{2\pi n q_\Delta}\mathbb{E}\left(\int_{-\pi m}^{\pi m} \frac{du}{|\widetilde{\phi_\Delta}(u)|^2}\right) \\
&\leq \frac{3\kappa}{2\pi n q_\Delta}\int_{-\pi m}^{\pi m} \frac{du}{|\phi_\Delta(u)|^2} + \frac{3\kappa}{2\pi n q_\Delta}\int_{-\pi m}^{\pi m} \mathbb{E}\left(\left|\frac{1}{\widetilde{\phi_\Delta}(u)} - \frac{1}{\phi_\Delta(u)}\right|^2\right) du \\
&\leq 3\kappa\, \text{pen}(m) + \frac{3\kappa}{2\pi n q_\Delta}\frac{2\pi m(1+2\mu\Delta)^4}{n} \leq 3\kappa\, \text{pen}(m) + \frac{c}{n}.
\end{aligned}$$

Using (28) and Lemma 2, we derive

$$\begin{aligned}
\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2 \mathbb{1}_\Omega) &\leq 3\|f - f_m\|^2 + 6\kappa\text{pen}(m) + \mathbb{E}\left([16p(m,\hat{m}) - 2\widehat{\text{pen}}(\hat{m})]\mathbb{1}_\Omega\right) + \frac{c}{n\Delta} \\
&\leq 3\|f - f_m\|^2 + 6(\kappa+8)\text{pen}(m) + 2\mathbb{E}\left([24\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m})]\mathbb{1}_\Omega\right) + \frac{c}{n\Delta}. \qquad (29)
\end{aligned}$$

Now we note that, as $1/q_\Delta \leq 2/\tilde{q}_\Delta$ on Ω,

$$\begin{aligned}
\mathbb{E}\left([24\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m})]\mathbb{1}_\Omega\right) &= \mathbb{E}\left[\left(\frac{24}{2\pi n q_\Delta}\int_{-\pi\hat{m}}^{\pi\hat{m}} \frac{du}{|\phi_\Delta(u)|^2} - \frac{\kappa}{2\pi n \tilde{q}_\Delta}\int_{-\pi\hat{m}}^{\pi\hat{m}} \frac{du}{|\widetilde{\phi_\Delta}(u)|^2}\right)\mathbb{1}_\Omega\right] \\
&\leq \mathbb{E}\left[\left(\frac{96}{2\pi n \tilde{q}_\Delta}\int_{-\pi\hat{m}}^{\pi\hat{m}} \frac{du}{|\widetilde{\phi_\Delta}(u)|^2} - \frac{\kappa}{2\pi n \tilde{q}_\Delta}\int_{-\pi\hat{m}}^{\pi\hat{m}} \frac{du}{|\widetilde{\phi_\Delta}(u)|^2}\right)\mathbb{1}_\Omega\right] \\
&\quad + \mathbb{E}\left[\left(\frac{96}{2\pi n \tilde{q}_\Delta}\int_{-\pi\hat{m}}^{\pi\hat{m}} \left|\frac{1}{\widetilde{\phi_\Delta}(u)} - \frac{1}{\phi_\Delta(u)}\right|^2 du\right)\mathbb{1}_\Omega\right].
\end{aligned}$$

Now we choose $\kappa \geq 96$ (which makes the first r.h.s. difference negative or zero), use that on Ω, $1/\tilde{q}_\Delta \leq (3/2)(1/q_\Delta)$ and that $\hat{m} \leq n\Delta$ which, together with Lemma 1 implies that $\mathbb{E}\left([24\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m})]\mathbb{1}_\Omega\right) \leq c/n$. Plugging this in (29) yields, for $\kappa \geq \kappa_0 = 96$ that, $\forall m \in \mathcal{M}_n$,

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2 \mathbb{1}_\Omega) \leq 3\|f - f_m\|^2 + 6(\kappa+8)\text{pen}(m) + \frac{c}{n\Delta}.$$

On the other hand

$$\begin{aligned}
\mathbb{P}(\Omega^c) &= \mathbb{P}\left(\left|\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\right| > \frac{1}{2q_\Delta}\right) \leq (2q_\Delta)^8\, \mathbb{E}\left(\left|\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta}\right|^8\right) \\
&\leq (2q_\Delta)^8 C'(4,\Delta)\frac{1}{\Delta^{12}n^4} \leq \frac{c}{(n\Delta)^4},
\end{aligned}$$

where the last line follows from Proposition 1. Moreover $\|\hat{f}_{\hat{m}} - f\|^2 = \|\hat{f}_{\hat{m}} - f_{\hat{m}}\|^2 + \|f - f_{\hat{m}}\|^2$. Now, $\|f - f_{\hat{m}}\|^2 \leq \|f\|^2$ and as $\|\hat{f}_{\hat{m}} - f_{\hat{m}}\|^2 \leq \|\hat{f}_{n\Delta} - f_{n\Delta}\|^2 \leq cn^3$, we obtain that

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2 \mathbb{1}_{\Omega^c}) \leq \frac{c}{n\Delta^4} := c'_\Delta/n.$$

This, together with (29) implies the result given in Theorem 1. □

5.3. **Proofs of Section 3.** The following bound is obtained as is Lemma 1:

$$\mathbb{E}\left(\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|^{2p}\right)\leq c_p\inf\left((n\Delta)^{-p}|H(v)|^{-4p},|H(v)|^{-2p}\right). \tag{30}$$

We state a useful Lemma:

Lemma 3. *Assume that* **[B]** (i)-(ii)(p) *hold. Then recalling that* $\psi=iG_\Delta/H_\Delta$ *(see (4)), we have*

$$\mathbb{E}\left(\left|\widetilde{\psi}(v)-\psi(v)\right|^{2p}\right)\leq\frac{\kappa_p}{(n\Delta)^p}\frac{1+|\psi(v)|^{2p}}{|H_\Delta(v)|^{2p}}.$$

Proof of Lemma 3. We omit the index Δ for simplicity. We have

$$\widetilde{\psi}-\psi=i\left(\frac{\hat{G}}{\tilde{H}}-\frac{G}{H}\right)=i(\hat{G}-G)\left(\frac{1}{\tilde{H}}-\frac{1}{H}\right)+i\frac{\hat{G}-G}{H}+iG\left(\frac{1}{\tilde{H}}-\frac{1}{H}\right) \tag{31}$$

so that a bound on (31) follows from bounding $\mathbb{E}(|\hat{G}(v)-G(v)|^{2p})$ and $\mathbb{E}\left(|\tilde{H}^{-1}(v)-H^{-1}(v)|^{2p}\right)$ for $p=1,2$. Clearly

$$\mathbb{E}(|\hat{G}(v)-G(v)|^2)=\frac{1}{n\Delta^2}\mathrm{Var}(Y_1(\Delta)e^{ivY_1(\Delta)})\leq\frac{\mathbb{E}(|Y_1(\Delta)|^2/\Delta)}{n\Delta}, \tag{32}$$

where $\mathbb{E}(|Y_1(\Delta)|^2/\Delta)=\mathbb{E}\Lambda\mathbb{E}\xi^2+\Delta\mathbb{E}(\Lambda^2)(\mathbb{E}(\xi))^2$. And for general p, the Rosenthal inequality yields

$$\begin{aligned}\mathbb{E}(|\hat{G}(v)-G(v)|^{2p}) &\leq \frac{C(2p)}{(n\Delta)^{2p}}\left\{n2^{2p}\mathbb{E}(|Y_1(\Delta)|^{2p})+[n\mathrm{Var}(Y_1(\Delta)e^{ivY_1(\Delta)})]^p\right\}\\ &\leq c(\frac{1}{(n\Delta)^{2p-1}}+\frac{1}{(n\Delta)^p})\end{aligned} \tag{33}$$

since **[B]**(ii)(p) holds and $n\Delta\geq 1$. Next the bound on $\mathbb{E}\left(|\tilde{H}^{-1}-H^{-1}|^{2p}\right)$ is given by (32). We conclude using (31), (32), (33) and (30). □

Proof of Proposition 4. Let

$$R(u)=\int_0^u\left(\frac{\hat{G}(v)}{\tilde{H}(v)}-\frac{G(v)}{H(v)}\right)dv.$$

Then to compute the risk of the estimator, we write:

$$\|\widetilde{f}_m-f\|^2=\|f-f_m\|^2+\|\widetilde{f}_m-f_m\|^2=\|f-f_m\|^2+\frac{1}{2\pi}\int_{-\pi m}^{\pi m}|\widetilde{f^*}(u)-f^*(u)|^2du$$

$$\begin{aligned}
|\widetilde{f^*}(u)-f^*(u)|^2 &\leq |\widetilde{f^*}(u)-f^*(u)|^2\mathbb{1}_{|R(u)|<1}+|\widetilde{f^*}(u)-f^*(u)|^2\mathbb{1}_{|R(u)|\geq 1}\\
&\leq |\widehat{f^*}(u)-f^*(u)|^2\mathbb{1}_{|R(u)|<1}+4\mathbb{1}_{|R(u)|\geq 1}\\
&\leq |f^*(u)|^2\left|\exp(R(u))-1\right|^2\mathbb{1}_{|R(u)|<1}+4\mathbb{1}_{|R(u)|\geq 1}\\
&\leq e^2|f^*(u)|^2|R(u)|^2\mathbb{1}_{|R(u)|\leq 1}+4|R(u)|\mathbb{1}_{|R(u)|>1}.
\end{aligned}$$

We prove now

$$\mathbb{E}(|R(u)|^2\mathbb{1}_{|R(u)|\leq 1})\leq \frac{c}{n\Delta}\left(M_1\int_0^{|u|}\frac{dv}{|H(v)|^2}+\left(\int_0^{|u|}\frac{|G(v)|}{|H(v)|^2}dv\right)^2\right), \tag{34}$$

$$\begin{aligned}
\mathbb{E}(|R(u)|\mathbb{1}_{|R(u)|>1}) &\leq c\left\{\left(M_p+\left(\int_0^{|u|}\left|\frac{G(v)}{H(v)}\right|^2dv\right)^p\right)\left(\frac{1}{n\Delta}\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^p\right.\\
&\qquad\left.+\frac{\mathbb{E}(|Y_1(\Delta)|^{2p}/\Delta)}{(n\Delta)^{2p-1}}\left(\int_0^{|u|}\frac{dv}{|H(v)|}\right)^{2p}\right\}
\end{aligned} \tag{35}$$

with $M_p=[\|G'\|_1^p+\mathbb{E}^{1/2}(Y_1^{2p}(\Delta)/\Delta)]$.

By decomposition (31), we write that $|R(u)|\leq R_1(u)+R_2(u)+R_3(u)$ with

$$R_1(u)=\left|\int_0^u\frac{\hat{G}(v)-G(v)}{H(v)}dv\right|,\quad R_2(u)=\left|\int_0^u G(v)\left(\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right)dv\right|$$

$$R_3(u)=\left|\int_0^u(\hat{G}(v)-G(v))\left(\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right)dv\right|.$$

Let $A_j:=\{|R(u)|\leq 1\}\cap\{R_j(u)=\max_{k\in\{1,2,3\}}R_k(u)\}$, then

$$\begin{aligned}
\mathbb{E}(|R(u)|^2\mathbb{1}_{|R(u)|\leq 1}) &\leq 9\mathbb{E}(R_1^2(u)\mathbb{1}_{A_1})+9\mathbb{E}(R_2^2(u)\mathbb{1}_{A_2})+\mathbb{E}(|R(u)|\mathbb{1}_{A_3})\\
&\leq 9\mathbb{E}(R_1^2(u))+9\mathbb{E}(R_2^2(u))+3\mathbb{E}(R_3(u)\mathbb{1}_{A_3}).
\end{aligned} \tag{36}$$

Then

$$\begin{aligned}
\mathbb{E}(R_1^2(u)) &\leq \frac{1}{n\Delta^2}\int_0^u\int_0^u\frac{\mathbb{E}(Y_1^2(\Delta)e^{i(v-w)Y_1(\Delta)})}{H(v)H(-w)}dvdw\\
&\leq \frac{1}{n\Delta}\left(\int_0^u\int_0^u\frac{1}{|H(v)|^2}|G'(v-w)|dvdw\right)^{1/2}\left(\int_0^u\int_0^u\frac{1}{|H(w)|^2}|G'(v-w)|dvdw\right)^{1/2}\\
&\leq \frac{\|G'\|_1}{n\Delta}\int_0^u\frac{dv}{|H(v)|^2}.
\end{aligned} \tag{37}$$

Moreover

$$\begin{aligned}
\mathbb{E}(R_2^2(u)) &\leq \int_0^u\int_0^u G(v)G(-w)\mathbb{E}\left[\left(\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right)\left(\frac{1}{\tilde{H}(-w)}-\frac{1}{H(-w)}\right)\right]dvdw\\
&\leq \frac{c}{n\Delta}\int_0^u\int_0^u\frac{|G(v)G(-w)|}{|H(v)|^2|H(w)|^2}dvdw=\frac{c}{n\Delta}\left(\int_0^u\frac{|G(v)|}{|H(v)|^2}dv\right)^2.
\end{aligned} \tag{38}$$

$$\text{Lastly} \quad \mathbb{E}(R_3(u)) \leq \int_0^u \mathbb{E}^{1/2}(|\hat{G}(v)-G(v)|^2)\mathbb{E}^{1/2}\left(\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|^2\right)dv$$

$$\leq c\frac{\mathbb{E}^{1/2}(Y_1^2(\Delta)/\Delta)}{n\Delta}\int_0^u \frac{1}{|H(v)|^2}dv \tag{39}$$

We plug (37)-(39) in (36) and we obtain (34). Let now $B_j := \{|R(u)| > 1\} \cap \{R_j(u) = \max_{k\in\{1,2,3\}} R_k(u)\}$. On B_j, $|R(u)| \leq 3R_j(u)$ and thus $3R_j(u) > 1$. Then

$$\begin{aligned}\mathbb{E}(|R(u)|\mathbb{1}_{|R(u)|>1}) &\leq 3(\mathbb{E}(R_1(u)\mathbb{1}_{B_1})+\mathbb{E}(R_2(u)\mathbb{1}_{B_2})+\mathbb{E}(R_3(u)\mathbb{1}_{B_3}))\\ &\leq 9^p(\mathbb{E}(R_1^{2p}(u))+\mathbb{E}(R_2^{2p}(u)))+3^p\mathbb{E}(R_3^p(u)\mathbb{1}_{B_3}).\end{aligned} \tag{40}$$

By applying Rosenthal's inequality and using the bound obtained in (37), we get

$$\mathbb{E}(R_1^{2p}(u)) \leq c\left(\|G'\|_1^p\left(\frac{1}{n\Delta}\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^p + \frac{\mathbb{E}(|Y_1(\Delta)|^{2p}/\Delta)}{(n\Delta)^{2p-1}}\left(\int_0^{|u|}\frac{dv}{|H(v)|}\right)^{2p}\right).$$

For R_2 we write

$$\mathbb{E}(R_2^{2p}(u)) \leq \left(\int_0^{|u|}\left(\frac{|G(v)|}{|H(v)|}\right)^2 dv\right)^p \mathbb{E}\left[\left(\int_0^{|u|}|H(v)|^2\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|^2 dv\right)^p\right].$$

Now we apply the Hölder inequality and inequality (30),

$$\begin{aligned}\mathbb{E}(R_2^{2p}(u)) &\leq \left(\int_0^{|u|}\left(\frac{|G(v)|}{|H(v)|}\right)^2 dv\right)^p\left(\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^{p-1}\int_0^{|u|}\frac{|H(v)|^{4p}}{|H(v)|^2}\mathbb{E}\left(\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|^{2p}\right)dv\\ &\leq c\left(\int_0^{|u|}\left(\frac{|G(v)|}{|H(v)|}\right)^2 dv\right)^p\left(\frac{1}{n\Delta}\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^p.\end{aligned}$$

For R_3 we apply the Hölder inequality again, and then the Cauchy Schwarz inequality, (33) and (30), to obtain

$$\begin{aligned}\mathbb{E}(R_3^p(u)) &\leq \left(\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^{p-1}\int_0^{|u|}\frac{|H(v)|^{2p}}{|H(v)|^2}\mathbb{E}\left(\left(|\hat{G}(v)-G(v)|\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|\right)^p\right)dv\\ &\leq \left(\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^{p-1}\int_0^{|u|}\frac{|H(v)|^{2p}}{|H(v)|^2}\mathbb{E}^{1/2}\left(|\hat{G}(v)-G(v)|^{2p}\right)\mathbb{E}^{1/2}\left(\left|\frac{1}{\tilde{H}(v)}-\frac{1}{H(v)}\right|^{2p}\right)dv\\ &\leq c\,\mathbb{E}^{1/2}(|Y_1(\Delta)|^{2p}/\Delta)\left(\frac{1}{n\Delta}\int_0^{|u|}\frac{dv}{|H(v)|^2}\right)^p.\end{aligned}$$

Plugging the three bounds in (40) gives (35). □

5.4. **Auxiliary result.**

Lemma 4. *Consider c, s nonnegative real numbers, and γ a real such that $2\gamma > -1$ if $c = 0$ or $s = 0$. Then, for all $m > 0$, $\int_{-m}^{m}(x^2+1)^{\gamma}\exp(c|x|^s)dx \approx m^{2\gamma+1-s}e^{cm^s}$, and if in addition $2\gamma > 1$ if $c = 0$ or $s = 0$, $\int_{m}^{\infty}(x^2+1)^{-\gamma}\exp(-c|x|^s)dx \approx m^{-2\gamma+1-s}e^{-cm^s}$.*

The proof of this lemma is based on integration by parts and is omitted. See also Lemma 2 p. 35 in Butucea and Tsybakov (2008a).

References

[1] Antonio, K. and Beirlant, J. (2007) Actuarial statistics with generalized linear mixed models. *Insurance Math. Econom.* **40**, 58-76.

[2] Belomestny, D. (2011) Statistical inference for time-changed Lévy processes via composite characteristic function estimation. *Ann. Statist.* **39**, 2205-2242.

[3] Birgé, L. and Massart, P. (2007) Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33-73.

[4] Bogsted, M. and Pitts, S M. (2010) Decompounding random sums: a nonparametric approach. *Ann. Inst. Statist. Math.* **62**, 855-872.

[5] Butucea, C. and Tsybakov, A. B. (2008a) Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52**, 24-39.

[6] Butucea, C. and Tsybakov, A. B. (2008b) Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52**, 237-249.

[7] Chesneau, C., Comte, F. and Navarro, F. (2013) Fast nonparametric estimation for convolutions of densities. *the Canadian Journal of Statistics*, **41**, 617-636.

[8] Comte, F., Duval, C. and Genon-Catalot, V. (2014) Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika* **77**, 163-183.

[9] 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.

[10] Davidian, M., and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data*, New York: Chapman and Hall.

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

[12] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling extremal events.* Berlin: Springer.

[13] van Es, B., Gugushvili, S. and Spreij, P. (2007) A kernel type nonparametric density estimator for decompounding. *Bernoulli* **13**, 672694.

[14] 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.

[15] Grandell, J. (1997) *Mixed Poisson processes.* Monographs on Statistics and Applied Probability, 77. Chapman & Hall, London.

[16] Gugushvili, S. (2012) Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, 282-307.

[17] Hall, P. and Heyde, C. C. (1980) *Martingale limit theory and its application.* Probability and Mathematical Statistics. Academic Press, New York-London.

[18] Mikosch, T. (2009) *Non-life insurance mathematics. An introduction with the Poisson process.* Second edition. Universitext. Springer-Verlag, Berlin.

[19] Neumann, M. H. (1997) On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* **7**, 307-320.

[20] Neumann, M. H. and Reiss, M. (2009) Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15**, 223-248.

[21] Pinheiro, J.C. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS.* New York, Springer.

[22] Tuckwell, H.C (1998). *Introduction to Theoretical Neurobiology: Volume 2.* Nonlinear and stochastic theories. Cambridge studies in biology 8. Cambridge University Press, New York.

[23] 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.

Corresponding author.
Fabienne COMTE
MAP5 UMR 8145, University Paris Descartes,
45, rue des Saints-Pères,
75006 PARIS, FRANCE
email. fabienne.comte@parisdescartes.fr

Supporting information

Additional supporting information may be found in the online version of this article at the publishers web site.

Appendix A. Additional proofs

A.1. Proof of Lemma 1

A.2. Proof of Proposition 1

A.3. Proof of Lemma 2

A.4. Proof of Proposition 3

A.5. Proof of Theorem 2

A.6. Proof of Proposition A.1

A.7. Talagrand inequality

Appendix B. Additional simulation results