

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

F. COMTE¹, C. DUVAL¹, V. GENON-CATALOT¹, AND J. KAPPUS²

ABSTRACT. Consider a mixed compound process $Y(t) = \sum_{i=1}^{N(\Lambda t)} \xi_i$ where N is a Poisson process with intensity 1, Λ a positive random variable, (ξ_i) a sequence of *i.i.d.* random variables with density f and $(N, \Lambda, (\xi_i))$ are independent. In this paper, we study nonparametric estimators of f by specific deconvolution methods. Assuming that Λ has exponential distribution with unknown expectation, we propose two types of estimators based on the observation of an *i.i.d.* sample $(Y_j(\Delta))_{1 \leq j \leq n}$ for Δ a given time. One strategy is for fixed Δ , the other for small Δ (with large $n\Delta$). Risks bounds and adaptive procedures are provided. Then, with no assumption on the distribution of Λ , we propose a nonparametric estimator of f based on the joint observation $(N_j(\Lambda_j \Delta), Y_j(\Delta))_{1 \leq j \leq n}$. Risks bounds are provided leading to unusual rates. The methods are implemented and compared via simulations. October 2, 2014

Keywords. Adaptive methods. Deconvolution. Mixed compound Poisson process. Nonparametric density estimation. Penalization method.

AMS Classification. 62M09–62G07

1. INTRODUCTION

Compound Poisson processes are commonly used in many fields of applications, especially in queuing and risk theory (see *e.g.* Embrechts *et al.* (1997), Grandell (1997), Mikosch (2009)). Nonparametric estimation of the jump size density in compound Poisson processes has been the subject of several recent contributions. 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 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 is random. The model is now as follows. Let Λ be a positive random variable, independent of $(N(t), t \geq 0)$

¹ Université Paris Descartes, MAP5, UMR CNRS 8145.

² Rostock Universität.

and of the sequence $(\xi_i, i \geq 1)$. Then,

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

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 are extremely popular in biostatistics and actuarial statistics. They are used in studies in which repeated measurements are taken on a series of individuals (see *e.g.* Davidian and Giltinan (1995), Pinheiro and Bates (2000), Antonio and Beirlant (2007), Belomestny (2011)). The randomness of parameters allows to account for the variability existing between subjects. This is why here, we assume that, for a given time Δ , we have *i.i.d.* observations $(Y_j(\Delta), j = 1, \dots, n)$ 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 estimating procedures based on *i.i.d.* data $(Y_j(\Delta), j = 1, \dots, n)$ may be compared with those of procedures based on increments $(X^\lambda(k\Delta) - X^\lambda((k-1)\Delta), k \leq n)$ for one trajectory. In particular, rates of convergence depend on $n\Delta$ for large n , and small or fixed Δ (not too large). Note that this topic corresponds to decompounding a random sum. Boegsted and Pitts (2010) assume that the distribution of the random number of terms is general but known and construct a nonparametric plug in estimator of f . In here, we assume that the distribution of Λ either contains unknown parameters or is completely unknown, so that the distribution $N(\Lambda t)$ is unknown.

The model we study belongs to the general class of random sums of random variables. This model is 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. The number of claims is possibly also observed. The mixed effect model to define the distribution of $M_j = M_j(\Delta)$ is also of common use (see Mikosch(2009) and Grandell (1997)). 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).

To fix notations, let $(\Lambda_j, j \geq 1)$ be *i.i.d.* with distribution $\nu(d\lambda)$, let $(N_j(t), j \geq 1)$ be *i.i.d.* Poisson processes with intensity 1 independent of $(\Lambda_j, j \geq 1)$ 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 $(\xi_i^j, j, i \geq 1)$ are *i.i.d.* with density f , and the sequence $(\xi_i^j, j, i \geq 1)$ is independent of $(\Lambda_j, (N_j(t)), j \geq 1)$. The paper is divided in two distinct parts, one is semi-parametric and the other purely nonparametric. 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 unobserved random intensities Λ_j 's have an exponential distribution with unknown parameter μ^{-1} (expectation μ). We define two different nonparametric estimators of f by a deconvolution approach. First, introducing

$$(2) \quad Q_\Delta(u) = \mathbb{E}(e^{iuY(\Delta)} \mathbf{1}_{Y(\Delta) \neq 0}), \quad \phi_\Delta(u) = \mathbb{E}e^{iuY(\Delta)},$$

we observe that the Fourier transform $f^*(u)$ of f satisfies $f^*(u) = Q_\Delta(u)/(q_\Delta \phi_\Delta(u))$, where $q_\Delta := \mathbb{P}(Y(\Delta) \neq 0)$. 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 $Q_\Delta(u), \phi_\Delta(u), q_\Delta$ where

we have to deal with the fact that $q_\Delta, \phi_\Delta(u)$ appear in the denominator of $f^*(u)$. Then, by Fourier inversion, we build a collection of nonparametric estimators $\hat{f}_m(x)$ of f associated with a cut off parameter m . Proposition 2.2 gives the bound of the \mathbb{L}^2 -risk of the estimator with fixed cut off parameter. Afterwards, we propose a data driven selection \hat{m} of m and prove that the corresponding estimator $\hat{f}_{\hat{m}}$ is adaptive (Theorem 2.1). The risk bounds are non asymptotic. 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 which performs well for small Δ . The idea is that for $\mu\Delta < 1$, the following series development holds:

$$(3) \quad f^*(u) = \sum_{k \geq 0} (-1)^k (1 + \mu\Delta)(\mu\Delta)^k (g_\Delta^*(u))^{k+1},$$

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, based on estimators of g_Δ and μ . For these two methods, no moment condition on f is imposed, thus fat tails density can be considered.

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 focus on estimating f without using any estimator for the distribution of Λ . 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 is based on the following idea. Assuming that $f^*(u) \neq 0$ for all u , $\mathbb{E}(\Lambda + |\xi|) < +\infty$ and $K_\Delta(u) := \mathbb{E}(\Lambda e^{iuY(\Delta)}) \neq 0$, we check that

$$(4) \quad \psi(u) := \frac{(f^*(u))'}{f^*(u)} = i \frac{G_\Delta(u)}{H_\Delta(u)},$$

where

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

Therefore, $\psi(u)$ can be estimated by using empirical counterparts of $G_\Delta(u), H_\Delta(u)$. As $f^*(0) = 1$, $f^*(u) = \exp(\int_0^u \psi(v)dv)$ can be estimated replacing ψ by an estimator. Afterwards, we proceed by deconvolution to define a collection of estimators \tilde{f}_m depending on a cut off parameter m . Proposition 3.1 gives the bound of the \mathbb{L}^2 -risk of \tilde{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 optimal 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 is simply 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 the first method of Section 2 performs well for all values of Δ except very small contrary to the second one, as expected from theoretical results. The methods of Section 3 perform well for all Δ and various distributions for Λ , including the exponential. A complete discussion on numerical results is given. Section 5 contains proofs. In the Appendix, auxiliary results needed in proofs are recalled. The last appendix contains supplementary material.

2. SEMI-PARAMETRIC STRATEGY 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:

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

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

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

Noting that

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

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

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

The following properties hold.

Proposition 2.1. *Under the Assumption:*

(H1) 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}_Δ , $1/\tilde{q}_\Delta$ 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

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

where $C(p, \mu_1)$ only depends on p, 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, Δ .

Note that Assumption **(H1)** is needed but definition (9) does require the knowledge of μ_0, μ_1 .

2.2. Notation. The following notations are used below. For $u : \mathbb{R} \rightarrow \mathbb{C}$ 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

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

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 Δ . 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

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

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

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

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:

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

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

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

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

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

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

Proposition 2.2. *Assume that Λ is $\mathcal{E}(\mu^{-1})$, $f \in \mathbb{L}^2(\mathbb{R})$ and that **(H1)** 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 $f_m^* = f^* \mathbf{1}_{[-\pi m, \pi m]}$ and where the constant c does not depend on n nor Δ .

We can see that 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 $\mathcal{S}(\alpha, L) = \{f \in \mathbb{L}^2(\mathbb{R}), \int |f^*(u)|^2 (1 + u^2)^\alpha du \leq L\}$,

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

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 $\mathcal{M}_n = \{1, \dots, n\Delta\}$, we select

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

Then we can prove the following result

Theorem 2.1. *Assume that Λ is $\mathcal{E}(\mu^{-1})$, $f \in \mathbb{L}^2(\mathbb{R})$ and that assumption **(H1)** holds. Then for $\kappa \geq \kappa_0 = 96$, we have*

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

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

The bounds of Proposition 2.2 and Theorem 2.1 are nonasymptotic and hold for all n and Δ . However, if Δ gets too small, the method deteriorates because q_Δ is small and $1/q_\Delta$ is badly estimated.

Remark 1. Theorem 2.1 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 $\kappa_0 = 1 + \eta$, $\eta > 0$, and explodes for $\kappa_0 = 1 - \eta$. To obtain 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. 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 $\mu\delta < 1$ and will use Proposition 2.1 above, and Formula (17) together with Proposition 2.3 below to estimate the terms of the series.

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

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

where $\alpha_m(\mu, \Delta)$, $m \geq 0$ is defined by (7). The conditional distribution of $Y(\Delta)$ given $Y(\Delta) \neq 0$ has density and Fourier transform given by (see (2), (8))

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

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

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

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

$$(16) \quad \|g_\Delta\| \leq \|f\|.$$

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 (12) and (9):

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

The following proposition, proved in A.2 and inspired by Chesneau *et al.* (2012), shows that we have estimators of convolutions of g_Δ with parametric rate as soon as $k \geq 2$:

Proposition 2.3. *If $f \in \mathbb{L}^2(\mathbb{R})$ and if (H1) holds,*

$$\widehat{g_{m,\Delta}^{*k}}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} (\widetilde{g}_\Delta^*(u))^k e^{-iux} du, \quad g_{m,\Delta}^{*k}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} (g_\Delta^*(u))^k e^{-iux} du.$$

Then, for all $k \geq 2$

$$\mathbb{E} \left(\|\widehat{g_{m,\Delta}^{*k}} - g_{m,\Delta}^{*k}\|^2 \right) \leq c(k, \mu_0, \mu_1) \left(\frac{m}{(n\Delta)^k} + \frac{D_k^2}{n\Delta} \|f\|^2 \right), \quad D_k = (3^k - 2^k - 1)/2.$$

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

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

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 2.4. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$, that (H1) holds, that Λ is $\mathcal{E}(\mu^{-1})$, and that $2\mu_1\Delta < 1$ and $\Delta < 1$. Then, for any $m \leq n\Delta$, we have*

$$\mathbb{E} \left(\|\hat{f}_{m,K} - f\|^2 \right) \leq \|f_m - f\|^2 + 12 \frac{(1 + \mu\Delta)^3}{\mu} \frac{m}{n\Delta} + 4A(\mu_1\Delta)^{2K+2} + \frac{E_K}{n\Delta},$$

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

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

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

To choose K in practice, we impose $\Delta^{2K+2} \leq 1/(n\Delta)$, so that the residual is negligible, that is

$$(18) \quad K \geq K_0 := \frac{1}{2} \left(\frac{\log(n)}{|\log(\Delta)|} - 3 \right),$$

even if this contradicts the fact that K is fixed (and thus independent on n).

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

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

We can prove

Theorem 2.2. *Assume that $f \in \mathbb{L}^2(\mathbb{R})$, that Λ is $\mathcal{E}(\mu^{-1})$, that **(H1)** holds and that $2\mu_1\Delta < 1$. Then there exists a numerical constant κ'_0 such that, for all $\kappa' \geq \kappa'_0$, we have*

$$\mathbb{E} \left(\|\hat{f}_{\hat{m}_K, K} - f\|^2 \right) \leq c \inf_{m \in \{1, \dots, [n\Delta]\}} \left(\|f_m - f\|^2 + \kappa' \frac{(1 + \mu\Delta)^3}{\mu} \frac{m}{n\Delta} \right) + 4A(\mu_1\Delta)^{2K+2} + \frac{E'_K}{n\Delta},$$

where c is a numerical constant, A is defined in Proposition 2.4 and E'_K is a constant depending on K , μ_0 , μ_1 and $\|f\|$.

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

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 $\text{bin}^-(a, 1/1 + \mu\Delta)$. We observe that $\mathbb{P}(Y(\Delta) = 0) = \left(\frac{1}{1+\mu\Delta} \right)^a$, so that we cannot estimate the two parameters a and μ . For instance if a is known, the two estimators of Sections 2.3 and 2.4 can be generalized, up to some additional but tedious computations.

3. NONPARAMETRIC STRATEGY

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 the previous section. The distribution of Λ only appears through $\mathbb{P}(Y_j(\Delta) = 0)$. Even in the case where the distribution of Λ is known up to two parameters, there occurs an identifiability problem.

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

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

and by derivation (see (5)),

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

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). Note that for $u \leq 0$, $\int_0^u = -\int_u^0$, and the formula is still valid. We deduce an estimator by setting

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

where

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

with

$$\widehat{f}^*(u) = \exp\left(\int_0^u \tilde{\psi}(v) dv\right), \quad \tilde{\psi}(v) = i \frac{\hat{G}_\Delta(v)}{\hat{H}_\Delta(v)},$$

$$\hat{G}_\Delta(v) = \frac{1}{n\Delta} \sum_{j=1}^n Y_j(\Delta) e^{ivY_j(\Delta)},$$

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

for some constant k .

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 $\forall u \in \mathbb{R}$, $|K_\Delta(u)| \geq K_0$.
(ii)(p) $\mathbb{E}(\xi^{2p}) < +\infty$, $\mathbb{E}(\Lambda^{2p}) < +\infty$.
(iii) $\|G'_\Delta\|_1 < +\infty$.

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

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

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

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

Thus, if $(f^*)'$ and $(f^*)''$ are integrable, then **[B]**(iii) holds. Then we can prove the following result.

Proposition 3.1. *Assume that **(H0)** and **[B]** hold. Let \tilde{f}_m be given by (20) and let Δ be fixed, $n\Delta \geq 1$. Then the following bound holds:*

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

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 $2p$.

The bounds are specific to our problem since the unknown function appears both in bias and variance terms.

3.2. Rate of the estimator. We study the resulting rate for the estimator 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.

Then $\|f - f_m\|^2 = O(m^{-2\alpha+1})$ so that $\alpha > 1/2$ is required for consistency of the estimator. For the variance terms, using the bound (21), we have

$$\begin{aligned}
\mathbb{V}_1 &:= \int_{-\pi m}^{\pi m} |f^*(u)|^2 \left(\int_0^{|u|} \frac{dv}{|f^*(v)|^2} + \left(\int_0^{|u|} \frac{|(f^*)'(v)|}{|f^*(v)|^2} dv \right)^2 \right) du = O(m^2), \\
\mathbb{V}_2 &:= \int_{-\pi m}^{\pi m} \left(1 + \left(\int_0^{|u|} \left| \frac{(f^*)'(v)}{f^*(v)} \right|^2 dv \right)^p \right) \left(\int_0^{|u|} \frac{dv}{|f^*(v)|^2} \right)^p du = O(m^{(2\alpha+1)p+1}),
\end{aligned}$$

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 possible, \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 6.2 recalled in the Appendix. Then $\|f - f_m\|^2 = O(m^{-1}e^{-(\pi m)^2})$, $\mathbb{V}_1 = O(m)$, $\mathbb{V}_2 = O(m^{2p-1}e^{p(\pi m)^2})$ 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 $\mathbb{V}_1/(n\Delta)$ yields 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 for the previous methods, we need to propose a data-driven selection of the cut off. As above, the bias is estimated up to a constant by $-\|\tilde{f}_m\|^2$. The penalty is built by estimating the variance term of the risk bound. Here, we have three terms and we do not know which term to choose to build a penalty. Thus select m as in the fixed Δ method for exponential Λ

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

with $\widehat{\text{pen}}(m)$ defined in (15) and is built using the $Y_j(\Delta)$'s only.

The numerical results confirm that this simple strategy works well.

3.4. Second estimator. A simple procedure is available for estimating f based on the joint observation $(N_j(\Lambda_j\Delta), Y_j(\Delta))_{1 \leq j \leq n}$. 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:

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

and

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

The following property holds.

Proposition 3.2. *Assume **(H0)**, $\mathbb{E}(\Lambda) < +\infty$, $\mathbb{E}(\Lambda e^{-\Lambda\Delta}) \geq k_0$ and $n\Delta \geq 1 \vee \frac{4k^2}{k_0^2}$. Then*

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

where f_m is such that $f_m^* = f^* \mathbf{1}_{[-\pi m, \pi m]}$

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 :

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

The proof of Proposition 3.2 follows the same lines as the analogous Proposition 2.2 and is omitted. The proof of adaptiveness of $\check{f}_{\tilde{m}}$ is also omitted.

The interest of this estimator is obviously its simplicity. However, it strongly depends on the observed value $\hat{\alpha}_1$ as the number of observations taken for the estimation is $n\hat{\alpha}_1$. If this value is too small, the estimator performs poorly.

4. ILLUSTRATIONS OF THE METHODS

In this section, we illustrate the estimators with data driven cut off on simulated data. 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 density on $[5, 6]$, a translated exponential $\mathcal{E}(2) + 5$, a translated *Beta* distribution $\text{Beta}(2, 2) + 5$. To compute \mathbb{L}_2 -risks, we use 1000 repetitions for $n\Delta = 200, 1000$ and 2500 repetitions for $n\Delta = 5000$.

The cut off m is selected among 200 equispaced values between 0.01 and 2. All methods require the calibration of the constant κ in penalties. After preliminary experiments, κ is taken equal to 0.21 in method 1 (first semi-parametric method), to 5 in method 2 (second semi-parametric method), and to 5 in the last method.

Figure 1 corresponds to a jump density mixture of Gaussian $0.4\mathcal{N}(-2, 1) + 0.6\mathcal{N}(3, 1)$ and Λ an exponential $\mathcal{E}(1)$. Figure 2 plots a Gumbel jump distribution with c.d.f. $F(x) = \exp(-\exp(-x))$, $x > 0$ and Λ either $\mathcal{E}(1)$ or $\mathcal{U}([1, 2])$. The truncation constant k is always taken equal to 0.5, except in (22) where $k = 0$.

We compare the two semiparametric methods for the Gaussian jump density and Λ exponential $\mathcal{E}(1)$. 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 2.4 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 value $\bar{n}_{\neq 0}$ of non zero data represents the number of data really used for estimation. 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 3, 4 in Section A). 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.

n	Method 1	Method 3	Simple
200	$5.6 \cdot 10^{-3}$ ($5.7 \cdot 10^{-3}$)	$4.9 \cdot 10^{-3}$ ($6.0 \cdot 10^{-3}$)	$5.9 \cdot 10^{-3}$ ($4.5 \cdot 10^{-3}$)
500	$2.7 \cdot 10^{-3}$ ($2.5 \cdot 10^{-3}$)	$3.6 \cdot 10^{-3}$ ($1.2 \cdot 10^{-3}$)	$2.6 \cdot 10^{-3}$ ($2.2 \cdot 10^{-3}$)
1000	$1.5 \cdot 10^{-3}$ ($1.7 \cdot 10^{-3}$)	$2.4 \cdot 10^{-3}$ ($5.2 \cdot 10^{-3}$)	$1.3 \cdot 10^{-3}$ ($9.2 \cdot 10^{-4}$)
2000	$7.9 \cdot 10^{-4}$ ($5.7 \cdot 10^{-4}$)	$2.0 \cdot 10^{-3}$ ($9.7 \cdot 10^{-3}$)	$7.2 \cdot 10^{-4}$ ($5.1 \cdot 10^{-4}$)
5000	$3.7 \cdot 10^{-4}$ ($2.9 \cdot 10^{-4}$)	$7.4 \cdot 10^{-4}$ ($9.4 \cdot 10^{-4}$)	$3.1 \cdot 10^{-4}$ ($2.2 \cdot 10^{-4}$)

TABLE 1. Mean of the \mathbb{L}_2 -risks for the semi-parametric method 1, the nonparametric method 3 and the simple method (see section 3.4). $\Delta = 1$, $\Lambda \sim \mathcal{E}(1)$ and f is $\mathcal{N}(0, 3)$; standard deviation in parenthesis.

In Table 1 we compare the first semi-parametric method to the nonparametric ones (method 3 and simple). They are comparable, with good results even for small values of n . The simple method 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

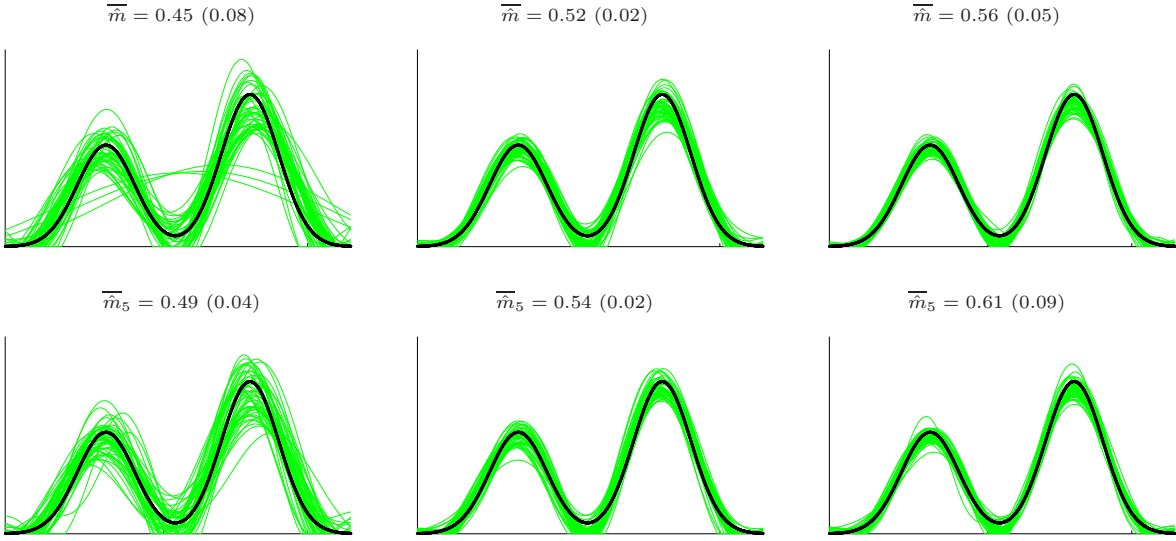


FIGURE 1. Estimation of f for a Gaussian mixture $0.4\mathcal{N}(-2, 1) + 0.6\mathcal{N}(3.1)$ for $n = 500$ (first column) $n = 2000$ (second column) and $n = 5000$ (third column) with the semiparametric method 1 (first line) and the semi parametric method 2 (second line, $K = 5$) for $\Delta = 1/2$. True density (bold black line) and 50 estimated curves (green lines).

theory nor in practice. The chosen distributions for Λ make the simple method perform worse than method 3. For $n = 200$, the risk of method 3 is twice better, for $n = 500$, it is three times better and for $n = 1000$, four times better. For larger n , the methods become equivalent. The simple method 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 estimating the distribution of Λ . In all cases, the values of m are small and stable.

Λ	n	Method 3	Simple
$\mathcal{U}([5, 6])$	200	$2.8 \cdot 10^{-2}$ ($6.6 \cdot 10^{-3}$)	$5.6 \cdot 10^{-2}$ ($1.5 \cdot 10^{-2}$)
	500	$9.0 \cdot 10^{-3}$ ($3.8 \cdot 10^{-3}$)	$2.8 \cdot 10^{-2}$ ($1.8 \cdot 10^{-2}$)
	1000	$3.3 \cdot 10^{-3}$ ($2.2 \cdot 10^{-3}$)	$1.3 \cdot 10^{-2}$ ($9.5 \cdot 10^{-3}$)
$\mathcal{E}(2) + 5$	200	$2.8 \cdot 10^{-2}$ ($6.8 \cdot 10^{-3}$)	$5.6 \cdot 10^{-2}$ ($1.7 \cdot 10^{-2}$)
	500	$9.0 \cdot 10^{-3}$ ($3.7 \cdot 10^{-3}$)	$2.7 \cdot 10^{-2}$ ($1.6 \cdot 10^{-2}$)
	1000	$3.4 \cdot 10^{-3}$ ($2.3 \cdot 10^{-3}$)	$1.3 \cdot 10^{-2}$ ($8.7 \cdot 10^{-3}$)
$\mathcal{Beta}(2, 2) + 5$	200	$2.7 \cdot 10^{-2}$ ($6.5 \cdot 10^{-3}$)	$5.6 \cdot 10^{-2}$ ($1.4 \cdot 10^{-2}$)
	500	$9.2 \cdot 10^{-3}$ ($4.0 \cdot 10^{-3}$)	$2.9 \cdot 10^{-2}$ ($1.7 \cdot 10^{-2}$)
	1000	$3.3 \cdot 10^{-3}$ ($2.2 \cdot 10^{-3}$)	$1.3 \cdot 10^{-2}$ ($8.9 \cdot 10^{-3}$)

TABLE 2. Mean of the L_2 -risks for the nonparametric method 3 and the simple method (see section 3.4); standard deviation in parenthesis. $\Delta = 1$, $\Lambda \sim \mathcal{U}([5, 6])$, $\mathcal{E}(2) + 5$, $\mathcal{Beta}(2, 2) + 5$ and f is $\mathcal{N}(0, 3)$.

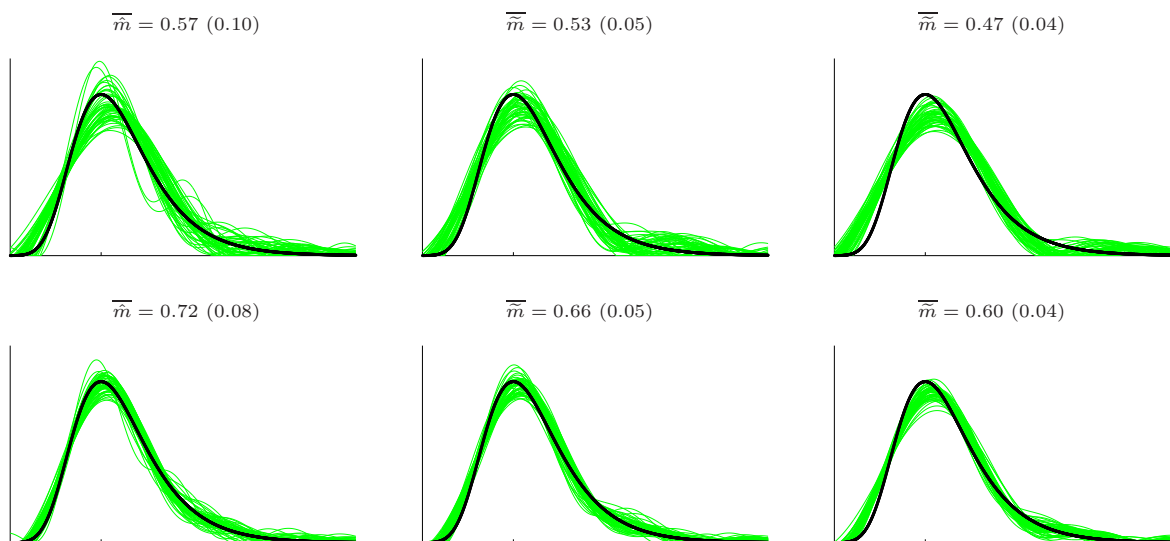


FIGURE 2. Estimation of f for a Gumbel distribution, for $n = 500$ (first line) $n = 2000$ (second line) with method 1 (first column) and method 3 (second and third column) for $\Delta = 1$. In the first two columns Λ is $\mathcal{E}(1)$ and in the third Λ is $\mathcal{U}([1, 2])$. True density (bold black line) and 50 estimated curves (green lines).

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. (1998) Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
- [4] Birgé, L. and Massart, P. (2007) Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**, 33-73.
- [5] Bogsted, M. and Pitts, S M. (2010) Decomposing random sums: a nonparametric approach. *Ann. Inst. Statist. Math.* **62**, 855-872.
- [6] Butucea, C. and Tsybakov, A. B. (2008a) Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52**, 24-39.
- [7] Butucea, C. and Tsybakov, A. B. (2008b) Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52**, 237-249.
- [8] Chesneau, C., Comte, F. and Navarro, F. (2013) Fast nonparametric estimation for convolutions of densities. *the Canadian Journal of Statistics*, **41**, 617-636.
- [9] Comte, F., Duval, C. and Genon-Catalot, V. (2014) Nonparametric density estimation in compound Poisson processes using convolution power estimators. *Metrika* **77**, 163-183.
- [10] 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.
- [11] Davidian, M., and Giltinan, D.M. (1995) *Nonlinear Models for Repeated Measurement Data*, New York: Chapman and Hall.
- [12] Duval, C. (2013) Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.* **123**, 3963-3986.
- [13] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling extremal events*. Berlin: Springer.

- [14] van Es, B., Gugushvili, S. and Spreij, P. (2007) A kernel type nonparametric density estimator for decomposing. *Bernoulli* **13**, 672694.
- [15] 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.
- [16] Grandell, J. (1997) *Mixed Poisson processes*. Monographs on Statistics and Applied Probability, 77. Chapman & Hall, London.
- [17] Gugushvili, S. (2012) Nonparametric inference for discretely sampled Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **48**, 282-307.
- [18] Hall, P. and Heyde, C. C. (1980) *Martingale limit theory and its application*. Probability and Mathematical Statistics. Academic Press, New York-London.
- [19] Klein, T. and Rio, E. (2005) Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.
- [20] Mikosch, T. (2009) *Non-life insurance mathematics. An introduction with the Poisson process*. Second edition. Universitext. Springer-Verlag, Berlin.
- [21] Neumann, M. H. (1997) On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* **7**, 307-320.
- [22] Neumann, M. H. and Reiss, M. (2009) Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15**, 223-248.
- [23] Pinheiro, J.C. and Bates, D.M. (2000) *Mixed-Effects Models in S and S-PLUS*. New York, Springer.
- [24] Tuckwell, H.C (1998). *Introduction to Theoretical Neurobiology: Volume 2*. Nonlinear and stochastic theories. Cambridge studies in biology 8. Cambridge University Press, New York.
- [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.

5. PROOFS

First we state a useful Lemma proved in Section A.1, similar to a Lemma stated in Neumann (1997).

Lemma 5.1. $\forall u \in \mathbb{R}$, $\mathbb{E}(|\frac{1}{\phi_\Delta(u)} - \frac{1}{\phi_\Delta(u)}|^{2p}) \leq c_p \left(\frac{1}{|\phi_\Delta(u)|^{2p}} \wedge \frac{n^{-p}}{|\phi_\Delta(u)|^{4p}} \right)$.

5.1. Proof of Proposition 2.1. By the Rosenthal Inequality (see Hall and Heyde (1980) p.23), we get

$$\mathbb{E}(|\hat{q}_\Delta - q_\Delta|^{2p}) \leq \frac{C(2p)}{n^{2p}} (nq_\Delta + (nq_\Delta(1 - q_\Delta))^p).$$

Now, $nq_\Delta = n\Delta\mu/(1 + \mu\Delta) \leq n\Delta\mu_1$ gives, under $n\Delta \geq 1$, the first bound:

$$\mathbb{E}(|\hat{q}_\Delta - q_\Delta|^{2p}) \leq C(2p)(\mu_1 + \mu_1^p) \left(\frac{\Delta}{n} \right)^p.$$

Now we write that

$$\frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta} = \left(\frac{1}{\hat{q}_\Delta} - \frac{1}{q_\Delta} \right) \mathbf{1}_{\hat{q}_\Delta > k\sqrt{\Delta/n}} - \frac{1}{q_\Delta} \mathbf{1}_{\hat{q}_\Delta \leq k\sqrt{\Delta/n}}.$$

With a proof analogous to Lemma 5.1, we get

$$\mathbb{E} \left(\left| \frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta} \right|^{2p} \right) \leq c \left(\frac{1}{q_\Delta^{2p}} \wedge \frac{(n/\Delta)^{-p}}{q_\Delta^{4p}} \right) = O\left(\frac{1}{(n\Delta^3)^p} \right).$$

For $\tilde{\mu}$, the decomposition is the following

$$\begin{aligned} \tilde{\mu} - \mu &= \frac{1}{\Delta} \left\{ \frac{\hat{q}_\Delta - q_\Delta}{1 - \hat{q}_\Delta} \mathbf{1}_{1 - \hat{q}_\Delta > k\sqrt{\Delta/n}} + q_\Delta \left(\frac{1}{1 - \hat{q}_\Delta} - \frac{1}{1 - q_\Delta} \right) \mathbf{1}_{1 - \hat{q}_\Delta > k\sqrt{\Delta/n}} \right. \\ &\quad \left. - \frac{q_\Delta}{1 - q_\Delta} \mathbf{1}_{1 - \hat{q}_\Delta \leq k\sqrt{\Delta/n}} \right\}, \end{aligned}$$

leading analogously to (10). \square

5.2. Proof of Proposition 2.2. 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 2.1 and Lemma 5.1.

Rosenthal's inequality implies for $p \geq 1$,

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

Then we have the decomposition, using Formula (11),

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

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

Then

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

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 2.1

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

and analogously by Lemma 5.1, $\mathbb{E} \left(\int_{-\pi m}^{\pi m} |T_2(u)|^2 du \right) \leq c' \|f\|^2 / n$. Then as $m \leq n\Delta$, with both Lemma 5.1 and Proposition 2.1, 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 2.2. \square

5.3. Proof of Theorem 2.1. Let $S_m = \{t \in \mathbb{L}^2(\mathbb{R}), t^* = t^* \mathbf{1}_{[-\pi m, \pi m]}\}$, 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

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

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

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

Writing that

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

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

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

Now, we write the decomposition

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

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

Clearly, the proof of Proposition 2.2, the Cauchy Schwarz inequality and $\|t^*\|^2 = 2\pi$ lead to $\mathbb{E} \left(\sup_{t \in S_{\hat{m} \vee m}, \|t\|=1} |R(t)|^2 \right) \leq c/(n\Delta)$. Thus, we have to study $\sup_{t \in S_{\hat{m} \vee m}, \|t\|=1} q_\Delta^{-1} \langle t^*, (\hat{Q}_\Delta - Q_\Delta)/\phi_\Delta \rangle$, for which, we can prove (see the proof in Section 5.4):

Lemma 5.2. *Under the Assumptions of Theorem 2.1, 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

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

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, $\forall m \in \mathcal{M}_n$,

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

Using (28) and Lemma 5.2, we derive

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

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

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

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 5.1 implies that $\mathbb{E}([24\text{pen}(\hat{m}) - \widehat{\text{pen}}(\hat{m})] \mathbf{I}_\Omega) \leq c/n$. Plugging this in (29) yields, for $\kappa \geq \kappa_0 = 96$ that, $\forall m \in \mathcal{M}_n$,

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

On the other hand

$$\begin{aligned} \mathbb{P}(\Omega^c) &= \mathbb{P} \left(\left| \frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta} \right| > \frac{1}{2} \frac{1}{q_\Delta} \right) \leq (2q_\Delta)^6 \mathbb{E} \left(\left| \frac{1}{\tilde{q}_\Delta} - \frac{1}{q_\Delta} \right|^6 \right) \\ &\leq (2q_\Delta)^6 C'(3, \Delta) \frac{1}{\Delta^9 n^3} \leq \frac{c}{(n\Delta)^3}, \end{aligned}$$

where the last line follows from Proposition 2.1. 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 $\|\hat{f}_{\hat{m}} - f_{\hat{m}}\|^2 \leq \|\hat{f}_{n\Delta} - f_{n\Delta}\|^2 \leq c(n\Delta)^2$, we obtain that

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2 \mathbf{I}_{\Omega^c}) \leq \frac{c}{n\Delta}.$$

This, together with (29) implies the result given in Theorem 2.1. \square

5.4. Proof of Lemma 5.2. For $t \in S_m$, t^* has support $I_m = [-\pi m, \pi m]$. To get the result, we apply the Talagrand inequality (see Appendix) to

$$t \mapsto \nu_n(t) = \frac{1}{2\pi n} \sum_{j=1}^n \int_{\mathbb{R}} t^*(-u) \frac{e^{iuY_j(\Delta)} \mathbf{I}_{Y_j(\Delta) \neq 0} - Q_\Delta(u)}{q_\Delta \phi_\Delta(u)} du.$$

First, we have

$$\mathbb{E} \left(\sup_{t \in S_{m' \vee m}, \|t\|=1} |\nu_n(t)|^2 \right) \leq \frac{1}{2\pi n} \int_{I_{m \vee m'}} \frac{1}{q_\Delta |\phi_\Delta^2(u)|} du := H^2.$$

Note that as $1 \leq 1/|\phi_\Delta(u)| \leq 1 + 2\mu\Delta$,

$$\frac{1}{q_\Delta} \frac{m \vee m'}{n} \leq H^2 \leq \frac{(1 + 2\mu\Delta)^2 m \vee m'}{q_\Delta n}.$$

Second, using $|\phi_\Delta^{-1}(u)| \leq 1 + 2\mu\Delta$ and $f^*(u) = Q_\Delta(u)/(q_\Delta \phi_\Delta(u))$, we get

$$\begin{aligned} & \sup_{t \in S_{m' \vee m}, \|t\|=1} \text{Var} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \frac{t^*(-u) e^{iuY_1(\Delta)} \mathbf{1}_{Y_1(\Delta) \neq 0}}{q_\Delta \phi_\Delta(u)} du \right) \\ & \leq \frac{1}{4\pi^2} \sup_{t \in S_{m' \vee m}, \|t\|=1} \mathbb{E} \left(\int_{\mathbb{R}^2} \frac{t^*(-u) t^*(w) e^{i(u-w)Y_1(\Delta)} \mathbf{1}_{Y_1(\Delta) \neq 0}}{q_\Delta^2 \phi_\Delta(u) \phi_\Delta(-w)} dudw \right) \\ & \leq \frac{1}{2\pi} \left(\int_{I_{m \vee m'}} \frac{|Q_\Delta(u-w)|^2}{q_\Delta^2 |\phi_\Delta(u)|^2 |\phi_\Delta(w)|^2} dudw \right)^{1/2} \\ & = \frac{1}{2\pi} \left(\int_{I_{m \vee m'}} \frac{|Q_\Delta(u-w)|^2}{q_\Delta^2 |\phi_\Delta(u-w)|^2} \frac{1}{|\phi_\Delta(u)|^2} \frac{|\phi_\Delta(u-w)|^2}{|\phi_\Delta(w)|^2} dudw \right)^{1/2} \\ & \leq (1 + 2\mu_1\Delta) \|f\| \left(\frac{1}{2\pi} \int_{I_{m \vee m'}} \frac{du}{|\phi_\Delta^2(u)|} \right)^{1/2} := v. \end{aligned}$$

Note that $v \leq c\sqrt{m \vee m'}$. Lastly

$$\sup_{t \in S_{m' \vee m}, \|t\|=1} \sup_{y \in \mathbb{R}} \left| \int_{\mathbb{R}} \frac{t^*(-u) e^{iuy} \mathbf{1}_{y \neq 0}}{q_\Delta \phi_\Delta(u)} du \right| \leq \sqrt{2\pi} \left(\int_{I_{m \vee m'}} \frac{du}{|q_\Delta \phi_\Delta(u)|^2} \right)^{1/2} \leq c \frac{\sqrt{m \vee m'}}{\Delta} := M$$

Thus,

$$\frac{v}{n} \propto \frac{\sqrt{m \vee m'}}{n}, \quad \frac{nH^2}{v} \propto \frac{\sqrt{m \vee m'}}{\Delta}, \quad \frac{M^2}{n^2} \propto \frac{m \vee m'}{(n\Delta)^2}, \quad \frac{nH}{M} \propto \sqrt{n\Delta},$$

and the Talagrand inequality with $\epsilon^2 = \frac{1}{4}$ gives the result. \square

5.5. Proof of Proposition 2.4. Let $F_k(\mu\Delta) := (1 + \mu\Delta)(\mu\Delta)^k$. We have

$$F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta) = (\tilde{\mu} - \mu)\Delta \left(\sum_{j=0}^{k-1} (\tilde{\mu}\Delta)^j (\mu\Delta)^{k-1-j} + \sum_{j=0}^k (\tilde{\mu}\Delta)^j (\mu\Delta)^{k-j} \right)$$

Then, we use inequality (10), which also implies $\mathbb{E}(|\tilde{\mu}|^{2p}) \leq C(p, \mu_1)$ to get

$$(30) \quad \mathbb{E}(F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta))^2 \leq C(k, \mu_1, \Delta) \frac{\Delta^{2k}}{n\Delta},$$

where the constant $C(k, \mu_1, \Delta) = C(k, \mu_1) + O(\Delta)$. We write

$$\hat{f}_{m,K} = \hat{f}_m^{(1)} + \hat{f}_{m,K}^{(2)}, \quad \hat{f}_m^{(1)} = (1 + \tilde{\mu}\Delta) \widehat{g_{m,\Delta}}.$$

We define

$$\tilde{f}_{m,K}^{(2)} = \sum_{k=1}^K (-1)^k F_k(\tilde{\mu}\Delta) g_{m,\Delta}^{*k+1}, \quad f_{m,K} = \sum_{k=0}^K (-1)^k F_k(\mu\Delta) g_{m,\Delta}^{*k+1},$$

and

$$f_m^{(1)} = (1 + \mu\Delta)g_{m,\Delta}, \quad f_{m,K}^{(2)} = f_{m,K} - f_m^{(1)}.$$

By the triangle inequality we have

$$\begin{aligned} \mathbb{E}(\|\hat{f}_{m,K} - f\|^2) &\leq 4 \left(\mathbb{E}(\|\hat{f}_m^{(1)} - f_m^{(1)}\|^2) + \mathbb{E}(\|\hat{f}_{m,K}^{(2)} - \tilde{f}_{m,K}^{(2)}\|^2) + \mathbb{E}(\|\tilde{f}_{m,K}^{(2)} - f_{m,K}^{(2)}\|^2) \right. \\ &\quad \left. + \|f_{m,K} - f_m\|^2 + \|f_m - f\|^2 \right) \\ &:= 4(\mathbb{E}(T_1 + T_2 + T_3) + T_4) + \|f_m - f\|^2, \end{aligned}$$

where $\|f_m - f\|^2$ is the usual bias term.

Now we successively study the terms T_i , for $i = 1, \dots, 4$.

Let us start by the study of T_1 . We split it again $T_1 \leq 3(T_{1,1} + T_{1,2} + T_{1,3})$ with

$$\begin{aligned} T_{1,1} &= \frac{1}{2\pi} \left(\frac{1 + \mu\Delta}{q_\Delta} \right)^2 \int_{-\pi m}^{\pi m} |\hat{Q}_\Delta(u) - Q_\Delta(u)|^2 du, \quad T_{1,2} = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{1 + \tilde{\mu}\Delta}{\tilde{q}_\Delta} - \frac{1 + \mu\Delta}{q_\Delta} \right|^2 |Q_\Delta(u)|^2 du, \\ T_{1,3} &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{1 + \tilde{\mu}\Delta}{\tilde{q}_\Delta} - \frac{1 + \mu\Delta}{q_\Delta} \right|^2 |\hat{Q}_\Delta(u) - Q_\Delta(u)|^2 du. \end{aligned}$$

As by (24), $\mathbb{E}(|\hat{Q}_\Delta(u) - Q_\Delta(u)|^2) \leq q_\Delta/n$, we find

$$\mathbb{E}(T_{1,1}) \leq \frac{(1 + \mu\Delta)^2}{q_\Delta} \frac{m}{n} = \frac{(1 + \mu\Delta)^3}{\mu} \frac{m}{n\Delta}.$$

This term is the main one. Now we prove that the two others are of order less than $O(1/(n\Delta))$.

It follows from Proposition 2.1 that

$$(31) \quad \mathbb{E} \left(\left| \frac{1 + \tilde{\mu}\Delta}{\tilde{q}_\Delta} - \frac{1 + \mu\Delta}{q_\Delta} \right|^{2k} \right) \leq C(k, \mu_0, \mu_1) \frac{1}{\Delta^{2k}} \frac{1}{(n\Delta)^k}.$$

Thus, using (31) with $k = 1$, and that $|Q_\Delta(u)| \leq q_\Delta |f^*(u)|$, we get

$$\mathbb{E}(T_{1,2}) \leq c_1(\mu_0, \mu_1) \frac{(q_\Delta/\Delta)^2}{n\Delta} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f^*(u)|^2 du \leq \frac{c(\mu_0, \mu_1) \|f\|^2}{n\Delta}.$$

Moreover, using (24), we have $\mathbb{E} \left(|\hat{Q}_\Delta(u) - Q_\Delta(u)|^4 \right) \leq c(2, \mu_1) (\Delta/n)^2$. Therefore, by the Cauchy-Schwarz Inequality and assuming that $n\Delta \geq 1$ and $m \leq n\Delta$, we obtain that

$$\mathbb{E}(T_{1,3}) \leq c(\mu_0, \mu_1) \frac{m}{(n\Delta)^2} \leq \frac{c(\mu_0, \mu_1)}{n\Delta}.$$

Gathering the terms, we have

$$\mathbb{E}(T_1) \leq 3 \frac{(1 + \mu\Delta)^3}{\mu} \frac{m}{n\Delta} + \frac{c(\mu_0, \mu_1, \|f\|)}{n\Delta}.$$

Next we study T_2 . By Inequality (30), as $1/(n\Delta) \leq 1$, we get

$$\begin{aligned} \mathbb{E}(\|\hat{f}_{m,K}^{(2)} - \tilde{f}_{m,K}^{(2)}\|^2) &= \mathbb{E}\left(\left\|\sum_{k=1}^K (-1)^k (F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta) + F_k(\mu\Delta)) (\widehat{g_{m,\Delta}^{*k+1}} - g_{m,\Delta}^{*k+1})\right\|^2\right) \\ &\leq c \sum_{k=1}^K \Delta^{2k} \mathbb{E}\left(\|\widehat{g_{m,\Delta}^{*k+1}} - g_{m,\Delta}^{*k+1}\|^2\right), \end{aligned}$$

where c depends on K and μ_1 . From Proposition 2.3, as $m \leq n\Delta$, this term is bounded by

$$\sum_{k=1}^K \Delta^{2k} c(k+1, \mu_0, \mu_1) \left(\frac{m}{(n\Delta)^{k+1}} + \frac{D_{k+1}^2}{n\Delta} \|f\|^2 \right) \leq \frac{c'(K, \mu_0, \mu_1)}{n\Delta}.$$

Now, to bound T_3 , we write that, using that $|g_\Delta^*(u)| \leq 1 \wedge |f^*(u)|$, for all u ,

$$\begin{aligned} 2\pi \|\tilde{f}_{m,K}^{(2)} - f_{m,K}^{(2)}\|^2 &\leq \int_{-\pi m}^{\pi m} \left| \sum_{k=1}^K (-1)^k (F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta)) (g_\Delta^*(u))^{k+1} \right|^2 du \\ &\leq \|f^*\|^2 \left| \sum_{k=1}^K |F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta)| \right|^2. \end{aligned}$$

Then, with Inequality (30), we get

$$\mathbb{E}\left(\|\tilde{f}_{m,K}^{(2)} - f_{m,K}^{(2)}\|^2\right) \leq K \|f\|^2 \sum_{k=1}^K \mathbb{E}(|F_k(\tilde{\mu}\Delta) - F_k(\mu\Delta)|^2) \leq c(K, \mu_1) \frac{\|f\|^2}{n\Delta}.$$

Thus T_3 is bounded by a term of order $1/(n\Delta)$.

Next we turn to T_4 . We have, using that $|g_\Delta^*(u)| \leq 1 \wedge |f^*(u)|$ for all u ,

$$\begin{aligned} 2\pi \|f_{m,K} - f_m\|^2 &= \|f_{m,K}^* - f_m^*\|^2 \leq \int_{-\pi m}^{\pi m} \left| \sum_{k=K+1}^{\infty} (1 + \mu_1\Delta)(\mu_1\Delta)^{k+1} (g_\Delta^*(u))^{k+1} \right|^2 du \\ &\leq \|f^*\|^2 \left(\sum_{k=K+1}^{\infty} (1 + \mu_1\Delta)(\mu_1\Delta)^{k+1} \right)^2 \\ &= \frac{(1 + \mu_1\Delta)^2}{(1 - \mu_1\Delta)^2} (\mu_1\Delta)^{2K+2} \|f^*\|^2 \leq 2\pi A (\mu_1\Delta)^{2K+2}, \end{aligned}$$

where A is defined in Proposition 2.4. It follows that,

$$\mathbb{E}(\|\hat{f}_{m,K} - f\|^2) \leq \|f_m - f\|^2 + 4 \left(A (\mu_1\Delta)^{2K+2} + 3 \frac{(1 + \mu\Delta)^3}{\mu} \frac{m}{n\Delta} + E_K \frac{1}{n\Delta} \right),$$

which is the result of Proposition 2.1. \square

5.6. Proof of Theorem 2.2. The only term involved in the adaptation is $T_{1,3}$, all the others are bounded independently of m . Thus the above bounds can be used, and $T_{1,3}$ is treated by using Talagrand Inequality applied to the underlying empirical process. The principle is as in Comte *et al.* (2014), proof of Theorem 4.1, with the use of the set Ω as in the proof of Theorem 2.1. For sake of conciseness, the proof is omitted.

5.7. Proofs of Section 3. We start by stating a useful Lemma.

Lemma 5.3. *Assume that [B] (i)-(ii)(p) hold. Then recalling that $\psi = iG_\Delta/H_\Delta$ (see (4)), we have*

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

Proof of Lemma 5.3. We omit the index Δ for simplicity. We have

$$(32) \quad \tilde{\psi} - \psi = i \left(\frac{\hat{G}}{\hat{H}} - \frac{G}{H} \right) = i(\hat{G} - G) \left(\frac{1}{\hat{H}} - \frac{1}{H} \right) + i \frac{\hat{G} - G}{H} + iG \left(\frac{1}{\hat{H}} - \frac{1}{H} \right)$$

so that a bound on (32) follows from bounding $\mathbb{E}(|\hat{G}(v) - G(v)|^{2p})$ and $\mathbb{E}\left(|\hat{H}^{-1}(v) - H^{-1}(v)|^{2p}\right)$ for $p = 1, 2$.

Clearly

$$(33) \quad \mathbb{E}(|\hat{G}(v) - G(v)|^2) = \frac{1}{n\Delta^2} \text{Var}(Y_1(\Delta)e^{ivY_1(\Delta)}) \leq \frac{\mathbb{E}(|Y_1(\Delta)|^2/\Delta)}{n\Delta},$$

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

$$(34) \quad \begin{aligned} \mathbb{E}(|\hat{G}(v) - G(v)|^{2p}) &\leq \frac{C(2p)}{(n\Delta)^{2p}} \left\{ n2^{2p}\mathbb{E}(|Y_1(\Delta)|^{2p}) + [n\text{Var}(Y_1(\Delta)e^{ivY_1(\Delta)})]^p \right\} \\ &\leq c \left(\frac{1}{(n\Delta)^{2p-1}} + \frac{1}{(n\Delta)^p} \right) \end{aligned}$$

since [B](ii)(p) holds and $n\Delta \geq 1$. Next the bound on $\mathbb{E}\left(|\hat{H}^{-1} - H^{-1}|^{2p}\right)$ is the following:

$$(35) \quad \mathbb{E}\left(\left|\frac{1}{\hat{H}(v)} - \frac{1}{H(v)}\right|^{2p}\right) \leq c_p \inf\left((n\Delta)^{-p}|H(v)|^{-4p}, |H(v)|^{-2p}\right).$$

The proof of (35) is similar to the proof of Lemma 5.1 and thus is omitted. We conclude using (32), (33), (34) and (35). \square

Proof of Proposition 3.1. Let

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

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

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

and

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

So we prove

$$(36) \quad \mathbb{E}(|R(u)|^2 \mathbf{1}_{|R(u)| \leq 1}) \leq \frac{c}{n\Delta} \left(M_1 \int_0^{|u|} \frac{dv}{|H(v)|^2} + \left(\int_0^{|u|} \frac{|G(v)|}{|H(v)|^2} dv \right)^2 \right)$$

and

$$(37) \quad \mathbb{E}(|R(u)| \mathbf{1}_{|R(u)| > 1}) \leq c \left\{ \left(M_p + \left(\int_0^{|u|} \left| \frac{G(v)}{H(v)} \right|^2 dv \right)^p \right) \left(\frac{1}{n\Delta} \int_0^{|u|} \frac{dv}{|H(v)|^2} \right)^p + \frac{\mathbb{E}(|Y_1(\Delta)|^{2p}/\Delta)}{(n\Delta)^{2p-1}} \left(\int_0^{|u|} \frac{dv}{|H(v)|^2} \right)^{2p} \right\}$$

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

By decomposition (32), we write that $|R(u)| \leq R_1(u) + R_2(u) + R_3(u)$ with

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

and

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

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

$$(38) \quad \begin{aligned} \mathbb{E}(|R(u)|^2 \mathbf{1}_{|R(u)| \leq 1}) &\leq 9\mathbb{E}(R_1^2(u) \mathbf{1}_{A_1}) + 9\mathbb{E}(R_2^2(u) \mathbf{1}_{A_2}) + \mathbb{E}(|R(u)| \mathbf{1}_{A_3}) \\ &\leq 9\mathbb{E}(R_1^2(u)) + 9\mathbb{E}(R_2^2(u)) + 3\mathbb{E}(R_3(u) \mathbf{1}_{A_3}). \end{aligned}$$

Then

$$(39) \quad \begin{aligned} \mathbb{E}(R_1^2(u)) &\leq \frac{1}{n\Delta^2} \int_0^u \int_0^u \frac{\mathbb{E}(Y_1^2(\Delta) e^{i(v-w)Y_1(\Delta)})}{H(v)H(-w)} dv dw \\ &\leq \frac{1}{n\Delta} \left(\int_0^u \int_0^u \frac{1}{|H(v)|^2} |G'(v-w)| dv dw \right)^{1/2} \left(\int_0^u \int_0^u \frac{1}{|H(w)|^2} |G'(v-w)| dv dw \right)^{1/2} \\ &\leq \frac{\|G'\|_1}{n\Delta} \int_0^u \frac{dv}{|H(v)|^2}. \end{aligned}$$

Moreover

$$(40) \quad \begin{aligned} \mathbb{E}(R_2^2(u)) &\leq \int_0^u \int_0^u G(v)G(-w) \mathbb{E} \left[\left(\frac{1}{\tilde{H}(v)} - \frac{1}{H(v)} \right) \left(\frac{1}{\tilde{H}(-w)} - \frac{1}{H(-w)} \right) \right] dv dw \\ &\leq \frac{c}{n\Delta} \int_0^u \int_0^u \frac{|G(v)G(-w)|}{|H(v)|^2 |H(w)|^2} dv dw = \frac{c}{n\Delta} \left(\int_0^u \frac{|G(v)|}{|H(v)|^2} dv \right)^2. \end{aligned}$$

Lastly

$$(41) \quad \begin{aligned} \mathbb{E}(R_3(u)) &\leq \int_0^u \mathbb{E}^{1/2}(|\hat{G}(v) - G(v)|^2) \mathbb{E}^{1/2} \left(\left| \frac{1}{\tilde{H}(v)} - \frac{1}{H(v)} \right|^2 \right) dv \\ &\leq c \frac{\mathbb{E}^{1/2}(Y_1^2(\Delta)/\Delta)}{n\Delta} \int_0^u \frac{1}{|H(v)|^2} dv \end{aligned}$$

We plug (39)-(41) in (38) and we obtain (36).

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

$$(42) \quad \begin{aligned} \mathbb{E}(|R(u)| \mathbf{1}_{|R(u)| > 1}) &\leq 3(\mathbb{E}(R_1(u) \mathbf{1}_{B_1}) + \mathbb{E}(R_2(u) \mathbf{1}_{B_2}) + \mathbb{E}(R_3(u) \mathbf{1}_{B_3})) \\ &\leq 9^p (\mathbb{E}(R_1^{2p}(u)) + \mathbb{E}(R_2^{2p}(u))) + 3^p \mathbb{E}(R_3^p(u) \mathbf{1}_{B_3}). \end{aligned}$$

By applying Rosenthal's inequality and using the bound obtained in (39), 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

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

Now we apply the Hölder inequality and inequality (35),

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

For R_3 we apply the Hölder Inequality again, and then the Cauchy Schwarz Inequality, (34) and (35), to obtain

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

Plugging the three bounds in (42) gives (37). \square

6. APPENDIX

The Talagrand inequality. 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 6.1. (*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 $\epsilon^2 > 0$

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

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

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

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.

Lemma 6.2. *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,b).

APPENDIX A. SUPPLEMENTARY APPENDIX

A.1. Proof of Lemma 5.1. Below, c is a constant which may change from line to line.

Let $B_n = (|\widehat{\phi}_\Delta(u)| \geq k/\sqrt{n})$. We write

$$\frac{1}{\widetilde{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)} = A_1 + A_2$$

with

$$A_1 = \left(\frac{1}{\widehat{\phi}_\Delta(u)} - \frac{1}{\phi_\Delta(u)} \right) \mathbf{1}_{B_n}, \quad A_2 = -\frac{1}{\phi_\Delta(u)} \mathbf{1}_{B_n^c}.$$

We have:

$$|A_1| \leq \frac{\sqrt{n}|\widehat{\phi}_\Delta(u) - \phi_\Delta(u)|}{k|\phi_\Delta(u)|}.$$

By the Rosenthal inequality, for all $q \geq 1$, $\mathbb{E}|\widehat{\phi}_\Delta(u) - \phi_\Delta(u)|^{2q} \leq 3/n^q$. Therefore, $\mathbb{E}|A_1|^{2p} \leq (3/(k|\phi_\Delta(u)|)^{2p})$. Obviously, $\mathbb{E}|A_2|^{2p} \leq 1/|\phi_\Delta(u)|^{2p}$. Now, we proceed to find the other bound. We have $A_1 = A'_1 + A''_1$ with

$$A'_1 = \frac{(\widehat{\phi}_\Delta(u) - \phi_\Delta(u))^2}{\phi_\Delta^2(u)\widehat{\phi}_\Delta(u)} \mathbf{1}_{B_n}, \quad A''_1 = \frac{\phi_\Delta(u) - \widehat{\phi}_\Delta(u)}{\phi_\Delta^2(u)} \mathbf{1}_{B_n}.$$

We have:

$$\mathbb{E}|A'_1|^{2p} \leq \frac{1}{|\phi_\Delta(u)|^{4p}} \left(\frac{\sqrt{n}}{k} \right)^{2p} \frac{3}{n^{2p}} = \frac{3}{k^{2p}} \frac{n^{-p}}{|\phi_\Delta(u)|^{4p}}, \quad \mathbb{E}|A''_1|^{2p} \leq 3 \frac{n^{-p}}{|\phi_\Delta(u)|^{4p}}.$$

Therefore, we also have $\mathbb{E}|A_1|^{2p} \leq cn^{-p}/|\phi_\Delta(u)|^{4p}$. Now, we study the second inequality for the term A_2 . First note that

$$B_n^c \subset (|\widehat{\phi}_\Delta(u) - \phi_\Delta(u)| \geq |\phi_\Delta(u)| - k/\sqrt{n}).$$

Moreover, $|\phi_\Delta(u)| - k/\sqrt{n} > |\phi_\Delta(u)|/2 \iff |\phi_\Delta(u)| > 2k/\sqrt{n}$. Thus,

$$\begin{aligned} \mathbb{P}(B_n^c) &\leq \mathbb{P}\left(|\widehat{\phi}_\Delta(u) - \phi_\Delta(u)| \geq \frac{|\phi_\Delta(u)|}{2}\right) + \mathbf{1}_{(|\phi_\Delta(u)| \leq 2k/\sqrt{n})} \\ &\leq \left(\frac{2}{|\phi_\Delta(u)|}\right)^{2p} \mathbb{E}|\widehat{\phi}_\Delta(u) - \phi_\Delta(u)|^{2p} + \mathbf{1}_{(|\phi_\Delta(u)|^{-1} \geq \sqrt{n}/2k)} \\ &\leq c \left(\frac{2}{|\phi_\Delta(u)|}\right)^{2p} n^{-p} + \left(\frac{2k}{\sqrt{n}}\right)^{2p} |\phi_\Delta(u)|^{-2p}. \end{aligned}$$

Thus $\mathbb{P}(B_n^c) \leq cn^{-p}/|\phi_\Delta(u)|^{2p}$. Finally, we also have:

$$(43) \quad \mathbb{E}|A_2|^{2p} \leq c \frac{n^{-p}}{|\phi_\Delta(u)|^{4p}}.$$

So the proof of Lemma 5.1 is complete. \square

A.2. Proof of Proposition 2.3. ,

We first prove that, for all $v \geq 1$ and $\mu \in [\mu_0, \mu_1]$ with $\mu_1 \Delta < 1$, we have

$$(44) \quad \sup_{u \in \mathbb{R}} \mathbb{E} \left(|\widetilde{g}_\Delta^*(u) - g_\Delta^*(u)|^{2v} \right) \leq \frac{C(v, \mu_0, \mu_1)}{(n\Delta)^v}.$$

We will use the following elementary inequality.

Lemma A.1. *Let $z = re^{i\theta}$ with $r \leq 1$, and $\hat{z} = \rho e^{i\omega}$, $\tilde{z} = e^{i\omega}$ with $\rho > 1$. Then $|\tilde{z} - z| \leq |\hat{z} - z|$.*

It follows from Lemma A.1 above that

$$|\widetilde{g}_\Delta^*(u) - g_\Delta^*(u)| \leq |\widehat{g}_\Delta^*(u) - g_\Delta^*(u)| = |T_1(u)| + |T_2(u)|$$

with

$$T_1(u) = \frac{1}{\widetilde{q}_\Delta}(\widehat{Q}_\Delta(u) - Q_\Delta(u)), \quad T_2(u) = Q_\Delta(u)\left(\frac{1}{\widetilde{q}_\Delta} - \frac{1}{q_\Delta}\right).$$

As $q_\Delta \leq \mu_1\Delta/(1 + \mu_1\Delta)$ and $\widetilde{q}_\Delta^{-1} \leq 2(\mu_0\Delta + 1)/(\mu_0\Delta)$, using Inequality (24), we get for all u ,

$$\mathbb{E}(|T_1(u)|^{2v}) \leq \frac{C(v, \mu_0, \mu_1)}{(n\Delta)^v}.$$

Second, as $|Q_\Delta(u)| \leq q_\Delta$, we obtain by Proposition 2.1, for all $v \geq 1$, and for all u ,

$$\mathbb{E}(|T_2(u)|^{2v}) \leq \left(\frac{q_\Delta}{\Delta}\right)^{2v} \frac{C(v, \mu_0, \mu_1)}{(n\Delta)^v}.$$

This gives (44).

Now we follow the lines of Chesneau *et al.* (2012). For $|v| \leq 1$ and $|w| \leq 1$, we have for any $k \geq 1$,

$$|w^k - v^k| = |(w - v)^k + \sum_{j=1}^{k-1} \binom{k}{j} v^j (w - v)^{k-j}| \leq |w - v|^k + |v| \sum_{j=1}^{k-1} \binom{k}{j} |w - v|^{k-j}$$

Thus, $|w^k - v^k|^2 \leq 2(|w - v|^{2k} + D_k^2 |v|^2 |w - v|^2)$. We apply this inequality for $w = \widetilde{g}_{m,\Delta}^*(u)$, $v = g_{m,\Delta}^*(u)$ and use (44) to obtain:

$$\mathbb{E}|(\widetilde{g}_{m,\Delta}^*(u))^k - (g_{m,\Delta}^*(u))^k|^2 \leq 2 \left(\frac{C(k, \mu_0, \mu_1)}{(n\Delta)^k} + D_k^2 |g_\Delta^*(u)|^2 \frac{C(1, \mu_0, \mu_1)}{n\Delta} \right).$$

Finally, the Plancherel theorem and Inequality (16) give

$$\begin{aligned} \mathbb{E} \left(\|\widehat{g}_{m,\Delta}^{*k} - g_{m,\Delta}^{*k}\|^2 \right) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \mathbb{E}(|(\widetilde{g}_{m,\Delta}^*(u))^k - (g_{m,\Delta}^*(u))^k|^2) du \\ &\leq c(k, \mu_0, \mu_1) \left(\frac{m}{(n\Delta)^k} + D_k^2 \frac{\|f\|^2}{n\Delta} \right). \end{aligned}$$

This ends the proof of Proposition 2.3. \square

A.3. Table of \mathbb{L}^2 -risks in semi-parametric methods 1 and 2. Because of computational limitations, we only make 100 repetitions (instead of 1000) when $\Delta = 0.01$ (first columns of Tables 3, 4).

Δ	0.01	0.1	0.5	0.9	1	2
n	20000	2000	400	220	200	100
$\bar{n}_{\neq 0}$	200	181	133	104	99	66
L_2 Risk	0.14 ($5.9 \cdot 10^{-3}$)	$6.2 \cdot 10^{-3}$ ($2.8 \cdot 10^{-3}$)	$3.9 \cdot 10^{-3}$ ($2.6 \cdot 10^{-3}$)	$5.3 \cdot 10^{-3}$ ($4.8 \cdot 10^{-3}$)	$5.6 \cdot 10^{-3}$ ($5.7 \cdot 10^{-3}$)	$2.1 \cdot 10^{-2}$ (0.06)
\bar{m}	0.01	0.25	0.29	0.29	0.29	0.29
n	100000	10000	2000	1110	1000	500
$\bar{n}_{\neq 0}$	990	908	666	526	499	333
L_2 Risk	0.014 (0.002)	$1.1 \cdot 10^{-3}$ ($5.0 \cdot 10^{-4}$)	$9.1 \cdot 10^{-4}$ ($5.8 \cdot 10^{-4}$)	$1.4 \cdot 10^{-3}$ ($1.6 \cdot 10^{-3}$)	$1.5 \cdot 10^{-3}$ ($1.7 \cdot 10^{-3}$)	$3.5 \cdot 10^{-3}$ ($5.0 \cdot 10^{-3}$)
\bar{m}	0.21	0.34	0.37	0.36	0.36	0.33
n	-	50000	10000	5550	5000	2500
$\bar{n}_{\neq 0}$	-	4545	3333	2969	2500	1666
L_2 Risk	-	$2.4 \cdot 10^{-4}$ ($9.5 \cdot 10^{-5}$)	$2.2 \cdot 10^{-4}$ ($1.3 \cdot 10^{-4}$)	$3.6 \cdot 10^{-4}$ ($2.3 \cdot 10^{-4}$)	$3.7 \cdot 10^{-4}$ ($2.9 \cdot 10^{-4}$)	$9.2 \cdot 10^{-4}$ ($1.1 \cdot 10^{-3}$)
\bar{m}	-	0.24	0.43	0.42	0.42	0.39

TABLE 3. Mean of the \mathbb{L}_2 -risks for the semi-parametric method 1 with $\Lambda \sim \mathcal{E}(1)$ (standard deviation in parenthesis) and f is $\mathcal{N}(0, 3)$; $\bar{n}_{\neq 0}$ is the mean of nonzero data; \bar{m} is the mean of selected \hat{m} 's.

Δ	0.01	0.1	0.5	0.9	1
n	20000	2000	400	220	200
$\bar{n}_{\neq 0}$	197	183	132	104	100
K	1	1	3	25	50
L_2 Risk	$1.6 \cdot 10^{-3}$ ($1.3 \cdot 10^{-3}$)	$1.9 \cdot 10^{-3}$ ($1.3 \cdot 10^{-3}$)	$3.8 \cdot 10^{-3}$ ($2.9 \cdot 10^{-3}$)	$2.6 \cdot 10^{-2}$ ($4.7 \cdot 10^{-2}$)	$1.5 \cdot 10^{16}$ ($5.0 \cdot 10^{17}$)
\bar{m}_K	0.36	0.35	0.30	0.25	0.24
n	100000	10000	2000	1110	1000
$\bar{n}_{\neq 0}$	992	910	667	525	499
K	1	1	4	32	50
L_2 Risk	$4.4 \cdot 10^{-4}$ ($3.5 \cdot 10^{-2}$)	$4.6 \cdot 10^{-4}$ ($3.1 \cdot 10^{-4}$)	$9.4 \cdot 10^{-4}$ ($6.6 \cdot 10^{-4}$)	$5.4 \cdot 10^{-3}$ ($3.5 \cdot 10^{-2}$)	$2.3 \cdot 10^3$ ($3.1 \cdot 10^4$)
\bar{m}_K	0.44	0.42	0.38	0.35	0.34

TABLE 4. Mean of the \mathbb{L}_2 -risks for the semi-parametric method 2 with $\Lambda \sim \mathcal{E}(1)$ (standard deviation in parenthesis) and f is $\mathcal{N}(0, 3)$; $\bar{n}_{\neq 0}$ is the mean of nonzero data; \bar{m} is the mean of selected \hat{m}_K 's.