An adaptive procedure for Fourier estimators: illustration to deconvolution and decompounding

Céline Duval, Johanna Kappus

An adaptive procedure for Fourier estimators: illustration to deconvolution and decompounding

Céline Duval* and Johanna Kappus†

Abstract

We introduce a new procedure to select the optimal cutoff parameter for Fourier density estimators that leads to adaptive rate optimal estimators, up to a logarithmic factor. This adaptive procedure applies for different inverse problems. We illustrate it on two classical examples: deconvolution and decompounding, i.e. non-parametric estimation of the jump density of a compound Poisson process from the observation of n increments of length $\Delta > 0$. For this latter example, we first build an estimator for which we provide an upper bound for its L^2-risk that is valid simultaneously for sampling rates Δ that can vanish, $\Delta := \Delta_n \to 0$, can be fixed, $\Delta_n \to \Delta_0 > 0$ or can get large, $\Delta_n \to \infty$ slowly. This last result is new and presents interest on its own. Then, we show that the adaptive procedure we present leads to an adaptive and rate optimal (up to a logarithmic factor) estimator of the jump density.

Keywords. Adaptive density estimation, deconvolution, decompounding, model selection
AMS Classification. 62C12, 62C20, 62G07.

1 Introduction

1.1 Motivation

In the literature on non-parametric statistics, and in particular in the literature dedicated to building minimax estimators, a lot of space is dedicated to adaptive procedures. Adaptivity may be understood as minimax-adaptivity, i.e. optimal rates of convergence are attained simultaneously over a collection of class of densities, for example, over a collection of Sobolev-balls. Adaptivity may also refer to proving non-asymptotic oracle bounds, i.e. having a procedure that mimics, up to a constant, the estimator that minimizes a given loss function. It is this last notion of adaptivity we adopt in this article.

Achieving adaptivity is a particular model selection issue for which there exist numerous techniques. Hereafter we mention some of them together with a non exhaustive list of references. Loosely speaking there exist three main approaches; thresholding techniques for wavelet density estimators (see e.g. [22, 23, 51]), penalized estimators (see e.g. [5, 1, 44, 42]) and pair wise comparison of estimators such as the Goldenshluger and Lepskii's procedure (see e.g.

*MAP5, UMR CNRS 8145, Université Paris Descartes.
†Institut für Mathematik, Universität Rostock.

[41, 40, 31, 29, 32, 30]). These techniques have been developed in a wide variety of contexts (for instance inverse problems or anisotropic multidimensional density estimation). Recently Lacour et al. [39] (see also the references therein) have introduced a new adaptive procedure for kernel density estimators, which is a modification of the Goldenshluger and Lepskii's method that has similar theoretical properties but is numerically more efficient. All the afore mentioned methods rely on the choice of an hyper parameter that needs to be calibrated, as explained in [39] numerical performances of the selected estimator are "very sensitive to this choice" and many studies have been devoted to the calibration of this hyper parameter (see e.g. Baudry et al. [2]). Another technique that is popular, especially in numerical studies, is cross validation, which does not rely on the calibration of an hyper parameter. However cross validation does not permit to obtain theoretical results in general, is time-consuming in practice and in some cases has very poor performances as explained in Delaigle and Gibels [19].

In the present paper, we put the stress on an adaptive method that was successfully applied in Duval and Kappus [26] and whose scope goes beyond the grouped data setting studied therein. The adaptive procedure presented below does not outperform existing techniques, it suffers a logarithmic loss, but has the advantage of being numerically simple and fast and permits to compute the optimal cutoff parameter for different classical inverse problems. This method also depends on the calibration of an hyper parameter, on the numerical study it seems that the method shows little sensibility to this parameter. To illustrate it both theoretically and numerically, we focus on two classical inverse problems: deconvolution, the density estimation problem is a particular case, and decompounding. These issues have raised a great interest in the literature (see the numerous references listed in Sections 2 and 3). For those two inverse problems we provide adaptive estimators that minimizes, up to a multiplicative constant, the upper bound which brings optimal rates of convergence.

In the deconvolution problem, one observes n i.i.d. realizations of $Y_i = X_i + \varepsilon_i$, where X_i and ε_i are independent, the density of ε_1 is known and the density f of X_1 is the quantity of interest. Our study relies on a well studied optimal Fourier estimator and for which the optimal upper bound for the L^2-risk is known. Starting from there, we make explicit our cutoff parameter in this context and show that its L^2-risk minimizes the optimal upper bound up to a logarithmic factor. We conduct an extensive simulation study which illustrates the stability of the procedure and we compare our results with a penalization procedure for which many results have been developed in this context.

In the decompounding problem, one discretely observes one trajectory of a compound Poisson process Z,

$$Z_t = \sum_{i=j}^{N_t} X_j, \quad t \geq 0,$$

where N is a Poisson process with intensity λ independent of the i.i.d. random variables $(X_j)_{j\in\mathbb{N}}$ with common density f. Let $\Delta > 0$, suppose we observe $(Z_{i\Delta},\ i = 1, \ldots, n)$, we aim at estimating f from these observations. The cases $\Delta \to 0$ (high frequency observations) or Δ being fixed (low frequency observations, often $\Delta = 1$) have been broadly studied in the literature (see the references given in Section 3) but were considered as separate cases. The case where Δ grows to infinity has never been studied. Therefore, we first study the problem

of estimating the jump density f of a compound Poisson process Z from $(Z_{i\Delta},\ i=1,\dots,n)$ for general sampling rates Δ. We invert the Lévy-Kintchine formula, relating the characteristic function of Z_Δ to the one of X_1. This approach is classical in the decompounding literature. We establish an upper bound for its L^2-risk where the dependency in Δ is made explicit. This upper bound is optimal simultaneously for $\Delta := \Delta_n \to \Delta_0 \in [0,\infty)$, it is also valid for $\Delta_n \to \infty$ under additional constraints (see Section 3). This result is new and presents interest on its own. The dependency in Δ_0 of the upper bound shows a deterioration as Δ_0 increases, which is expected. But, we identify regimes where $\Delta_n \to \infty$ such that $\Delta_n < 1/4\log(n\Delta_n)$ as $n \to \infty$ and where the estimator remains consistent, and presumably rate optimal. Heuristically, if the jump density has finite variance, for simplicity assume f is centered with unit variance, then, the law of each increment can be approximated as follows $X_\Delta = \sqrt{\lambda\Delta}\zeta_\Delta$, where $\zeta_\Delta \to \mathcal{N}(0,1)$ as $\Delta \to \infty$. Therefore, one would expect that in these regimes non-parametric estimation is impossible as each increment is close, in law, to a parametric Gaussian variable. When Δ goes too rapidly to infinity, namely as a power of $n\Delta_n$, Duval [25] shows that consistent non-parametric estimation of f is impossible, regardless of the choice of the loss function, by showing that it is always possible to build two different compound Poisson processes with different jump densities for which the statistical experiments generated by their increments are asymptotically equivalent. The results of the present paper complement the knowledge on decompounding. Finally, we show that our adaptive procedure leads to an adaptive and rate optimal estimator of the jump density f, up to a logarithmic loss, for all sampling rates such that $\Delta_n < 1/4\log(n\Delta_n)$ as $n \to \infty$, this condition is fulfilled for fixed or vanishing Δ.

The article is organized as follows. In the remaining of this Section we describe the idea of our adaptive procedure. In Section 2 we illustrate, both theoretically and numerically, its performances for the deconvolution problem. In the numerical part we compare our procedure with a penalized adaptive optimal estimator. Section 3 is dedicated to the decompounding problem. Finally, Section 5 gathers the proofs of the main results.

1.2 Methodology

Notations. We introduce some notations which are used throughout the rest of the text. Given a random variable Z, $\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$ denotes the characteristic function of Z. For $f \in \mathrm{L}^1(\mathbb{R})$, $\mathcal{F}f(u) = \int e^{iux} f(x)\,\mathrm{d}x$ is understood to be the Fourier transform of f. Moreover, we denote by $\|\cdot\|$ the L^2-norm of functions, $\|f\|^2 := \int |f(x)|^2\,\mathrm{d}x$. Given some function $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, we denote by f_m the uniquely defined function with Fourier transform $\mathcal{F}f_m = (\mathcal{F}f)\mathbb{1}_{[-m,m]}$.

Statistical setting. Consider n i.i.d. realizations Y_j, $j = 1,...,n$ of a random variable Y with Lebesgue-density f_Y. Suppose Y is related to a variable X, with Lebesgue-density f through a known transformation $\mathbf{T}$ relating their characteristic functions: $\varphi_Y = \mathbf{T}(\varphi_X)$, where $\mathbf{T}$ can be linear (e.g. deconvolution) or not (grouped data (see [46, 26]), decompounding). We are interested in estimating the density f of X from the (Y_j).

We focus on estimators based on Fourier methods, which is convenient for several classes of inverse problems as the convolution operation corresponds to multiplication. If the transformation $\mathbf{T}$ admits a continuous inverse, we build an estimator $\widehat{\varphi}_{X,n}$ of φ_X from the observations $(Y_1, \dots, Y_n)$,

$$\widehat{\varphi}_{X,n}(u) = \mathbf{T}^{-1}\big(\widehat{\varphi}_{Y,n}(u)\big) \quad \text{where } \widehat{\varphi}_{Y,n}(u) := \frac{1}{n}\sum_{j=1}^{n} e^{iuY_j},\ u \in \mathbb{R}.$$

Cutting off in the spectral domain and applying Fourier inversion gives an estimator of f

$$\widehat{f}_m(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\widehat{\varphi}_{X,n}(u)\,\mathrm{d}u, \quad \forall m > 0,\ x \in \mathbb{R}.$$

The performance of the estimator is measured with L^2-loss. The choice of the cutoff parameter is crucial: the goal is to select m that mimics the optimal cutoff $m^\star$ which minimizes the L^2-risk,

$$\mathbb{E}\big[\|\widehat{f}_{m^\star} - f\|^2\big] = \inf_{m\geq 0}\Big\{\frac{1}{2\pi}\int_{[-m,m]^c}|\varphi_X(u)|^2\,\mathrm{d}u + \frac{1}{2\pi}\int_{-m}^{m}\mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big]\,\mathrm{d}u\Big\}. \quad (1.1)$$

This optimal value m^* usually depends on the unknown regularity of f and is hence not feasible. We propose a procedure to select a random cutoff $\widehat{m}_n$, which can be calculated from the observations, and for which the L^2-risk is close to the one of $\widehat{f}_{m^*}$, meaning that we can establish an oracle bound

$$\mathbb{E}\big[\|\widehat{f}_{\widehat{m}_n} - f\|^2\big] \leq C\inf_{m\geq 0}\Big\{\frac{1}{2\pi}\int_{[-m,m]^c}|\varphi_X(u)|^2\,\mathrm{d}u + \frac{1}{2\pi}\int_{-m}^{m}\mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big]\,\mathrm{d}u\Big\} + r_n,$$

for a positive constant C and r_n a negligible remainder. We call $\widehat{f}_{\widehat{m}_n}$ adaptive rate optimal estimator of f.

Heuristic of the adaptive procedure. If $\mathbf{T}^{-1}$ is differentiable and if we can show that for some positive constant C, $\mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big] \leq C|(\mathbf{T}^{-1})'\big(\varphi_Y(u)\big)|^2\mathbb{E}\big[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2\big]$, we have,

$$\mathbb{E}\big[\|\widehat{f}_m - f\|^2\big] \leq \frac{1}{2\pi}\int_{[-m,m]^c}|\varphi_X(u)|^2\,\mathrm{d}u + \frac{C}{n}\int_{-m}^{m}\big|(\mathbf{T}^{-1})'\big(\varphi_Y(u)\big)\big|^2\,\mathrm{d}u, \quad m \geq 0. \quad (1.2)$$

The quantity $(\mathbf{T}^{-1})'$ is explicit in the grouped data setting (see [46, 26]), but also in the deconvolution and decompounding cases (see e.g. Sections 2 and 3 hereafter). The second term in the right hand side of the latter inequality is a majorant of the integrated variance of the estimator; using a majorant of the variance term is the starting point of many adaptive procedures such as penalized procedures or the Goldenshluger and Lepskii's procedure, where from this majorant, one tries to find a cutoff such that the empirical bias and the majorant are of the same order.

Denote by $\overline{m}_n$ the arginf of the right hand side of (1.2). If the upper bound (1.2) is optimal, meaning that it has the same order as (1.1), then asymptotically it holds that

$m^\star \asymp \overline{m}_n$. Differentiating in m the right hand side of (1.2) gives that the cutoff $\overline{m}_n$ that minimizes this quantity satisfies

$$|\varphi_X(\overline{m}_n)|^2 = \frac{C}{n}\big|(\mathbf{T}^{-1})'\big(\varphi_Y(\overline{m}_n)\big)\big|^2 \iff |\mathbf{T}(\varphi_Y(\overline{m}_n))|^2 = \frac{C}{n}\big|(\mathbf{T}^{-1})'\big(\varphi_Y(\overline{m}_n)\big)\big|^2. \tag{1.3}$$

Clearly, (1.3) has an empirical version and it is tempting to select $\widehat{m}_n$ accordingly. This inspires to select the cutoff parameter in the following ensemble,

$$\widehat{m}_n \in \Big\{m > 0, \Big|\frac{\mathbf{T}(\widehat{\varphi}_{Y,n}(m))}{(\mathbf{T}^{-1})'(\widehat{\varphi}_{Y,n}(m))}\Big| = \frac{1}{\sqrt{n}}\big(1+\kappa\sqrt{\log n}\big)\Big\} \wedge n, \tag{1.4}$$

for some $\kappa > 0$. However, the solution of (1.3) may not be unique; but considering the minimum (as in [26] where the density of interest also plays the role of the noise) or the maximum of this ensemble (as in the sequel), it is uniquely determined. Many adaptive procedures such as penalization methods minimizes an empirical version of the upper bound (1.2) whereas the spirit of (1.4) consists in finding the zeroes of an empirical version of the derivative in m of the upper bound (1.2). Roughly speaking, the difference between our procedure and a penalization procedure is the same as the difference between Z-estimators and M-estimators.

The quantity involved in the definition of $\widehat{m}_n$ also appears in the definition of the estimator. This explains why the procedure performs numerically fast. It has proven to be asymptotically minimax (up to a logarithmic loss) in the grouped data setting and we show that it also leads to adaptive rate optimal estimators (up to a logarithmic loss) in deconvolution and decompounding inverse problems.

2 Deconvolution

2.1 Statistical setting

Suppose that $X_1, \dots, X_n$ are i.i.d. with density f and are accessible through the noisy observations

$$Y_j = X_j + \varepsilon_j,\ j = 1, \dots, n.$$

Assume that the (ε_j) are i.i.d., independent of the (X_j) and such that $\forall u \in \mathbb{R},\ \varphi_\varepsilon(u) \neq 0$. Suppose that the distribution of ε_1 is known. This last assumption can be softened, the procedure allows a straightforward generalization to the case where the distribution of ε_1 can be estimated from an additional sample, see Neumann [47].

A deconvolution estimator of the characteristic function φ_X of X is given by

$$\frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}, \quad \text{with } \widehat{\varphi}_{Y,n}(u) := \frac{1}{n}\sum_{j=1}^{n} e^{iuY_j}, \quad u \in \mathbb{R},$$

denoting the empirical characteristic function. Since φ_X is a characteristic function, its absolute value is bounded by 1 and the estimator can hence be improved by using the definition

$$\widehat{\varphi}_{X,n}(u) := \frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)} \frac{1}{\max\{1, \big|\frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}\big|\}}, \quad u \in \mathbb{R}. \tag{2.1}$$

Cutting off in the spectral domain and applying Fourier inversion gives the estimator

$$\widehat{f}_m(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\widehat{\varphi}_{X,n}(u)\,\mathrm{d}u, \quad x \in \mathbb{R}. \tag{2.2}$$

This estimator and adaptation techniques have been extensively studied in the literature, including in more general settings than above. Optimal rates of convergence and adaptive procedures are well known if $d = 1$ (see e.g. [10, 52, 53, 28, 9, 7, 8, 49, 12] for L^2-loss functions or [43] for the L^∞-loss). Results have also been established for multivariate anisotropic densities (see e.g. [16] for L^2-loss functions or [50] for L^p-loss functions, $p \in [1, \infty]$). Deconvolution with unknown error distribution has also been studied (see e.g. [47, 20, 35, 45], if an additional error sample is available, or [17, 21, 36, 15, 38] under other set of assumptions).

2.2 Risk bounds and adaptive bandwidth selection

The following risk bound is well known in the literature on deconvolution.

$$\mathbb{E}\big[\|\widehat{f}_m - f\|^2\big] \le \|f - f_m\|^2 + \frac{1}{2\pi n}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} = \frac{1}{2\pi}\Big(\int_{[-m,m]^c}|\varphi_X(u)|^2\,\mathrm{d}u + \frac{1}{n}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}\Big). \tag{2.3}$$

This upper bound is the sum of a bias term that decreases with m and a variance term, increasing with m. We select $\overline{m}_n$, the optimal cutoff parameter, such that the upper bound (2.3) is minimal

$$\overline{m}_n \in \operatorname*{arginf}_{m\ge 0}\Big\{\int_{[-m,m]^c}|\varphi_X(u)|^2\,\mathrm{d}u + \frac{1}{n}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}\Big\}.$$

At $\overline{m}_n$ both terms in (2.3) are of the same order which gives the optimal cutoff. Differentiating the right hand side with respect to m, we find that the following holds for the optimal cutoff parameter:

$$|\varphi_X(\overline{m}_n)|^2 = \frac{1}{n|\varphi_\varepsilon(\overline{m}_n)|^2} \quad\Longleftrightarrow\quad |\varphi_X(\overline{m}_n)\varphi_\varepsilon(\overline{m}_n)| = |\varphi_Y(\overline{m}_n)| = \frac{1}{\sqrt{n}}. \tag{2.4}$$

This equality has an empirical version and we select $\widehat{m}_n$ accordingly. In order to ensure adaptivity the following heuristic consideration is helpful. When the characteristic function is replaced by its empirical version, the standard deviation is of the order $n^{-1/2}$. Consequently, estimating φ_Y by $\widehat{\varphi}_{Y,n}$ makes sense for $|\varphi_Y| \ge n^{-1/2}$. If $|\varphi_Y| < n^{-1/2}$, the noise is dominant so the estimator might be set to zero. This inspires to re-define the estimator of φ_Y as follows:

$$\widetilde{\varphi}_{Y,n}(u) = \widehat{\varphi}_{Y,n}(u)\mathbb{1}_{\{|\widehat{\varphi}_{Y,n}(u)|\ge\kappa_n n^{-1/2}\}}, \quad u \in \mathbb{R},$$

with the threshold value $\kappa_n := (1 + \kappa\sqrt{\log n})$. The constant $\kappa > 0$ is specified below. Then, the estimator of f given in (2.2) is modified using the following instead of (2.1)

$$\widetilde{\varphi}_{X,n}(u) := \frac{\widetilde{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}\frac{1}{\max\{1, \big|\frac{\widetilde{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}\big|\}}, \quad u \in \mathbb{R}$$

and

$$\widetilde{f}_m(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\widetilde{\varphi}_{X,n}(u)\,\mathrm{d}u, \quad x \in \mathbb{R}.$$

Note that the upper bound (2.3) remains valid for the estimator $\widetilde{f}_m$, thanks to the result:

Lemma 2.1. *Let $z = re^{i\theta}$, with $r \leq 1$, $\widehat{z} = \rho e^{iw}$, $\rho > 1$, and $\widetilde{z} = e^{iw}$. Then, $|\widetilde{z} - z| \leq |\widehat{z} - z|$.*

We define the empirical cutoff parameter $\widehat{m}_n$ as follows. Since $\widehat{\varphi}_{Y,n}$ may show an oscillatory behavior and the solution of (2.4) may not be unique, we consider

$$\widehat{m}_n = \max\left\{m > 0 : |\widehat{\varphi}_{Y,n}(m)| = \kappa_n n^{-1/2}\right\} \wedge n^{\alpha}, \tag{2.5}$$

for some $\alpha \in (0, 1]$. It is worth emphasizing that the calculation of $\widehat{m}_n$ does only rely on the empirical characteristic function $\widehat{\varphi}_{Y,n}$ and does not require the evaluation of penalty terms depending on the (perhaps unknown) φ_ε.

Theorem 2.1. *Let $\widehat{m}_n$ defined as in (2.5), with $\kappa > 0$ and $\alpha \in (0, 1]$. Then, there exist a positive constant C_1 depending only on the choice of κ and a universal positive constants C_2 such that*

$$\mathbb{E}[\|f - \widetilde{f}_{\widehat{m}_n}\|^2] \leq C_1 \inf_{m\in[0,n^\alpha]}\left\{\int_{[-m,m]^c} |\varphi_X(u)|^2\,\mathrm{d}u + \frac{\log n}{n}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}\right\} + C_2 n^{\alpha-\kappa^2/2}.$$

Theorem 2.1 is non asymptotic and ensures that the estimator $\widetilde{f}_{\widehat{m}_n}$ automatically reaches the bias-variance compromise, up to a logarithmic factor and the multiplicative constant C_1. The logarithmic loss is technical; it permits to control the deviation of the empirical characteristic functions from the true characteristic function by the Hoeffding inequality. Without changing the strategy of the proof removing the additional log factor seems difficult. Proof of Theorem 2.1 is self contained.

Regarding the choice of α and κ in (2.5), it is always possible to take $\alpha = 1$. Note that the case $\alpha > 1$ is not interesting as, even in the direct problem $\varepsilon = 0$ a.s., if $m > n$ the variance term in (2.3) no longer tends to 0. Taking $\alpha < 1$ is possible only if one has additional information on the target density f. For instance, if one knowns that f is in a Sobolev class of regularity β, for some $\beta \geq \beta_0 > 0$,

$$f \in \mathscr{S}(\beta, L) := \left\{f \in \mathbf{F},\ \int_{\mathbb{R}} (1 + |u|)^{2\beta}|\mathcal{F}f(u)|^2\,\mathrm{d}u \leq L\right\} \tag{2.6}$$

where $\mathbf{F}$ is the set of densities with respect to the Lebesgue measure. Then, it holds that $\|f - f_m\|^2 \asymp m^{-2\beta}$ and straightforward computations lead to $m^\star \lesssim \left(\frac{n}{\log n}\right)^{\frac{1}{2\beta+1}}$ (regardless the the asymptotic decay of φ_ε). Then, one may restrict the interval for $\widehat{m}_n$ to $[0, n^\alpha]$ where $1 > \alpha > \frac{1}{2\beta_0+1}$. Second, the choice of κ must be such that $n^{\alpha-\kappa^2/2}$ is negligible, the choice $\kappa > \sqrt{2}$ always works. The following numerical study suggests that the procedure is stable in the choice of κ.

2.3 Numerical results.

Stability of the procedure. To illustrate the performances of the method and the influence of the parameter κ we proceed as follows. For different densities f, namely, Uniform $\mathcal{U}[1,3]$, Gaussian $\mathcal{N}(2,1)$, Cauchy, Gamma $\Gamma(2,1)$ and the mixture $0.7\mathcal{N}(4,1)+0.3\Gamma(2,\frac{1}{2})$, and for different values of κ we compute the adaptive L^2 risks from $M=1000$ Monte Carlo iterations. The results are displayed on Figures 1, 2 and 3. We consider three different settings:

- The direct density estimation problem (Figure 1): we observe i.i.d. realizations of f. It is a particular deconvolution problem where $\varepsilon = 0$ a.s.
- Deconvolution problem with ordinary smooth noise (Figure 2): the error ε is Gamma $\Gamma(2,1)$ i.e. $|\varphi_\varepsilon|$ decays as $|u|^{-2}$ asymptotically.
- Deconvolution problem with super smooth noise (Figure 3): the error ε is Cauchy i.e. $|\varphi_\varepsilon|$ decays as $e^{-|u|}$ asymptotically.

On Figures 1, 2 and 3 we observe that the adaptive rates are small and that the procedure is stable on the choice of kappa. We observe on these three cases, that the value of κ should not be chosen too large but that for a wide range of values the performances are similar. In practice, the value of n is fixed and there is a natural boundary for κ, indeed observe that it is useless to increase κ if $(1+\kappa\log n)n^{-1/2}\geq 1$ as the selection rule (2.5) will be constant equal to n^α. Moreover, we expect that if $(1+\kappa\log n)n^{-1/2}$ gets too large, e.g. larger than $1/2$ the performances of the adaptive estimator should deteriorate. This practical consideration encourages to choose κ smaller than $(\sqrt{n}-1)\sqrt{\log(n)}^{-1}$. In Figures 1, 2 and 3 it appears that for all the meaningful values of κ, e.g. smaller than $\frac{1}{2}(\sqrt{n}-1)\sqrt{\log(n)}^{-1}$ for instance, the performances of the adaptive estimator are similar.

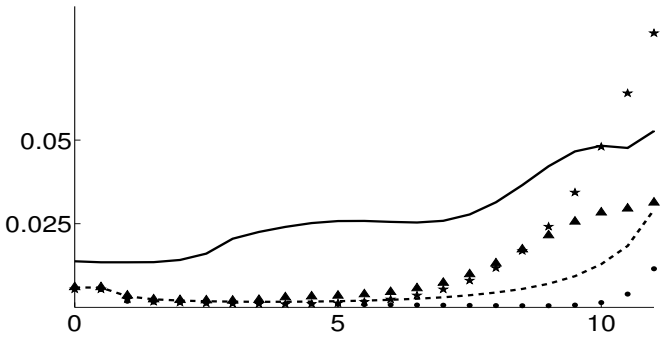

Figure 1: **Direct problem:** Computations by $M=1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa\leq(\sqrt{n}-1)\sqrt{\log(n)}^{-1}=11.6$ (x axis). Estimation of f from $n=1000$ i.i.d. direct realizations for different distributions: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1)+0.3\Gamma(2,\frac{1}{2})$ (triangles).

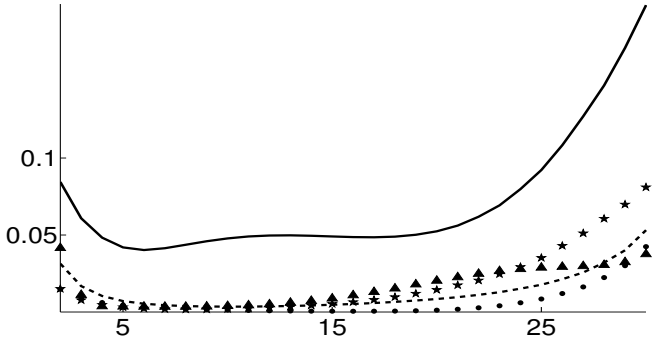

Figure 2: **Deconvolution problem (ordinary smooth case):** Computations by $M = 1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa \leq (\sqrt{n}-1)\sqrt{\log(n)}^{-1} = 32.6$ (x axis). Estimation of f from $n = 10000$ i.i.d. direct of $X + \varepsilon$ where ε has distribution $\Gamma(2,1)$ (i.e. $\varphi_\varepsilon(u) = (1-iu)^{-2}$) and for different distributions for X: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles).

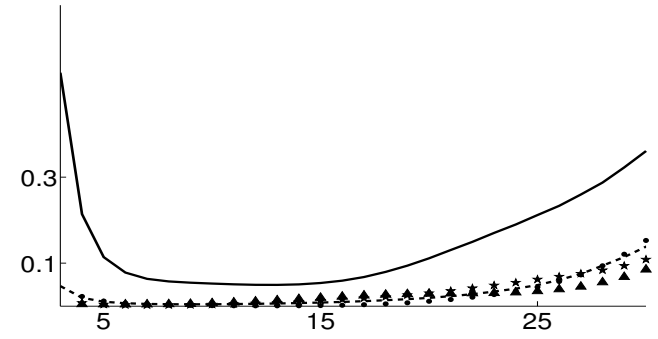

Figure 3: **Deconvolution problem (super smooth case):** Computations by $M = 1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa \leq (\sqrt{n}-1)\sqrt{\log(n)}^{-1} = 32.6$ (x axis). Estimation of f from $n = 10000$ i.i.d. realizations of $X + \varepsilon$ where ε has Cauchy distribution (i.e. $\varphi_\varepsilon(u) = e^{-|u|}$) and for different distributions for X: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles). For the uniform distribution, the rates where stable around the value 0.5, they do not appear on the Figure not to spoil the readability of the other curves.

Comparison with other procedures. We compare the performances of our procedure for $\kappa = 8$, with a penalization procedure and with an oracle. For the penalization procedure, we follow Comte and Lacour [17] and consider the adaptive estimator $\widehat{f}_{\widetilde{m}_n}$ which is the estimator defined in (2.2) where

$$\widetilde{m}_n = \underset{m \in [0, M_n]}{\operatorname{argmin}} \{-\|\widehat{f}_m\|^2 + \mathrm{pen}(m)\}, \ \mathrm{pen}(m) = K\Big(\frac{\Delta(m)}{\log(m+1)}\Big)^2 \frac{\Delta(m)}{n},$$

where $M_n > 0$, $K > 0$ and $\Delta(m) = \frac{1}{2\pi}\int_{[-m,m]} |\varphi_\varepsilon(u)|^{-2}\,\mathrm{d}u$, which is known in our setting. The parameter M_n is chosen as the maximal integer such that $1 \leq \frac{\Delta(m)}{n} \leq 2$. For the parameter K it is calibrated by preliminary simulation experiments. For calibration strategies (dimension jump and slope heuristics), the reader is referred to Baudry et al. [2]. Here, we test a grid of values of the Ks from the empirical error point of view, to make a relevant choice; the tests are conducted on a set of densities which are different from the one considered hereafter, to avoid overfitting. After these preliminary experiments, K is chosen equal to 2 which is the same value as the one considered in Comte and Lacour [17]. The standard errors are given in parenthesis. The running times for each risks of the penalization procedure and our procedure are similar; our procedure is barely faster. However, one should take into account that a preliminary calibration step seems obsolete in our case. In deconvolution problems, the theoretical optimal K can be in some cases far away from the practically optimal K and may vary with the sample size explaining the nessecity of this calibration step (see e.g. Kappus and Mabon [38] where the practical optimal value of K was much smaller than the value predicted by the theory).

Second, an oracle estimator is computed $\widehat{f}_{m^\star}$, which is the estimator defined in (2.2) where $m^\star$ corresponds to the following oracle bandwidth

$$m^\star = \underset{m>0}{\operatorname{argmin}}\ \mathbb{E}[\|f - \widehat{f}_m\|^2].$$

This oracle can be explicitly evaluated when f is known. We denote these different risks by R, for the risk of our procedure, R_{pen} for the penalized estimator and R_{or} for the oracle procedure. All these risks are computed on 1000 Monte Carlo iterations. The results are gathered in Tables 1 for the Gamma density, 2 for the mixture and 3 for the Cauchy density where $\mathcal{C}$ stands for the Cauchy distribution. In each case both an ordinary smooth and a super smooth errors are considered.

Comparaison of the different methods. Tables 1, 2 and 3 show that all the procedures behave as expected; the L^2-risks decreases with n and are smaller in the case of an ordinary smooth deconvolution problem than in the case of a super smooth deconvolution problem. The estimator with the smallest risk is the oracle, and the penalized risks are most of the time smaller than our procedure which is consistent with the fact that our procedure has a logarithmic loss and is asymptotic. More precisely for small values of n our procedure does not perform as well as the penalized method. But for larger values of n it is competitive. We can exhibit particular cases where our procedure is more stable in the choice of the hyper parameter than the penalized procedure, even on large sample sizes (see Figure 4 for example). This is due to the fact that the penalized constant K that is suitable for small values of n is different than for larger values of n. In practice a logarithmic term in n is added in the penalty term, that is theoretically unnecessary and entails a logarithmic loss but improves the numerical results. If we add this logarithmic term (we replace $K = 2$ with $\widetilde{K}\log(n)^{2.5}$ with $\widetilde{K} = 0.3$ and the multiplying $\log(n)^{2.5}$ factor as suggested in Comte et al. [18]). This second penalty procedure performs well for all values of n and when n gets large it has similar performances as our procedure (see Table 4). For our procedure, changing κ for smaller values of n does not improve the results.

f / f_ε	n	R	$\widehat{m}$	R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
		$\Gamma(2,1)$					
$\Gamma(2,1)$	500	4.31×10^{-2}	1.05	1.97×10^{-2}	0.80	0.74×10^{-2}	0.66
		(0.20)	(0.07)	(0.01)	(0.05)	(0.36×10^{-2})	(0.14)
	1000	1.74×10^{-2}	0.98	1.70×10^{-2}	0.94	0.59×10^{-2}	0.72
		(0.13)	(0.04)	(0.03)	(0.03)	(0.28×10^{-2})	(0.14)
	5000	0.40×10^{-2}	0.85	1.30×10^{-2}	1.32	0.31×10^{-2}	0.91
		(0.06)	(0.01)	(0.01)	(0.05)	(0.13×10^{-2})	(0.15)
$\mathcal{C}$	500	5.27×10^{-2}	0.90	1.21×10^{-2}	0.56	0.92×10^{-2}	0.61
		(0.23)	(0.07)	(0.69×10^{-2})	(0.04)	(0.48×10^{-2})	(0.12)
	1000	1.84×10^{-2}	0.84	0.98×10^{-2}	0.70	0.70×10^{-2}	0.67
		(0.13)	(0.04)	(0.61×10^{-2})	(0.03)	(0.34×10^{-2})	(0.13)
	5000	0.51×10^{-2}	0.70	0.71×10^{-2}	1.10	0.39×10^{-2}	0.82
		(0.07)	(0.01)	(0.01×10^{-2})	(0.02)	(0.17×10^{-2})	(0.13)

Table 1: Comparaison of the different adaptive estimators for the Gamma distribution.

f / f_ε	n	R	$\widehat{m}$	R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
		$0.7\mathcal{N}(4,1) + 0.3\Gamma(4, \frac{1}{2})$					
$\Gamma(2,1)$	500	1.77×10^{-2}	0.92	0.78×10^{-2}	0.78	0.33×10^{-2}	0.59
		(0.13)	(0.05)	(0.65×10^{-2})	(0.03)	(0.17×10^{-2})	(0.16)
	1000	0.74×10^{-2}	0.89	0.73×10^{-2}	0.87	0.26×10^{-2}	0.67
		(0.08)	(0.03)	(0.64×10^{-2})	(0.03)	(0.15×10^{-2})	(0.14)
	5000	0.13×10^{-2}	0.79	0.38×10^{-2}	1.10	0.01×10^{-2}	0.82
		(0.03)	(0.01)	(0.30×10^{-2})	(0.01)	(0.06×10^{-3})	(0.11)
$\mathcal{C}$	500	2.78×10^{-2}	0.82	0.73×10^{-2}	0.55	0.43×10^{-2}	0.50
		(0.15)	(0.05)	(0.50×10^{-2})	(0.05)	(0.20×10^{-2})	(0.14)
	1000	1.02×10^{-2}	0.77	0.65×10^{-2}	0.67	0.34×10^{-2}	0.58
		(0.10)	(0.03)	(0.52×10^{-2})	(0.03)	(0.16×10^{-2})	(0.15)
	5000	0.21×10^{-2}	0.66	0.59×10^{-2}	0.94	0.16×10^{-2}	0.74
		(0.04)	(0.01)	(0.48×10^{-2})	(0.02)	(0.10×10^{-2})	(0.11)

Table 2: Comparaison of the different adaptive estimators on a mixture.

f / f_ε	n	$\mathcal{C}$					
		R	$\widehat{m}$	R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
$\Gamma(2,1)$	500	2.69×10^{-2}	0.59	1.00×10^{-2}	0.68	0.67×10^{-2}	0.62
		(0.16)	(0.07)	(0.67×10^{-2})	(0.03)	(0.29×10^{-2})	(0.10)
	1000	1.00×10^{-2}	0.84	0.97×10^{-2}	0.82	0.49×10^{-2}	0.67
		(0.09)	(0.04)	(0.70×10^{-2})	(0.05)	(0.21×10^{-2})	(0.10)
	5000	0.27×10^{-2}	0.69	0.81×10^{-2}	1.18	0.21×10^{-2}	0.82
		(0.05)	(0.01)	(0.57×10^{-2})	(0.04)	(0.10×10^{-2})	(0.10)
$\mathcal{C}$	500	2.50×10^{-2}	0.74	1.12×10^{-2}	0.45	0.86×10^{-2}	0.56
		(0.16)	(0.07)	(0.27×10^{-2})	(0.01)	(0.34×10^{-2})	(0.09)
	1000	1.00×10^{-2}	0.68	0.74×10^{-2}	0.59	0.62×10^{-2}	0.62
		(0.10)	(0.04)	(0.32×10^{-2})	(0.03)	(0.24×10^{-2})	(0.09)
	5000	0.52×10^{-2}	0.54	0.73×10^{-2}	0.93	0.29×10^{-2}	0.74
		(0.07)	(0.01)	(0.52×10^{-2})	(0.03)	(0.11×10^{-2})	(0.09)

Table 3: Comparaison of the different adaptive estimators for the Cauchy distribution.

f / f_ε	n	$\mathcal{G}$		$\mathcal{M}$		$\mathcal{C}$	
		$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$	$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$	$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$
$\Gamma(2,1)$	500	1.17×10^{-2}	0.72	0.63×10^{-2}	0.69	0.83×10^{-2}	0.62
		(0.81×10^{-2})	(0.03)	(0.51×10^{-2})	(0.02)	(0.46×10^{-2})	(0.005)
	1000	0.94×10^{-2}	0.82	0.40×10^{-2}	0.75	0.59×10^{-2}	0.69
		(0.63×10^{-2})	(0.04)	(0.30×10^{-2})	(0.02)	(0.32×10^{-2})	(0.03)
	5000	0.62×10^{-2}	1.08	0.20×10^{-2}	0.94	0.31×10^{-2}	0.91
		(0.62×10^{-2})	(0.01)	(0.16×10^{-2})	(0.02)	(0.19×10^{-2})	(0.02)
$\mathcal{C}$	500	1.34×10^{-2}	0.45	0.60×10^{-2}	0.45	1.31×10^{-2}	0.41
		(0.43×10^{-2})	(0.01)	(0.25×10^{-2})	(0.01)	(0.24×10^{-2})	(0.02)
	1000	0.91×10^{-2}	0.59	0.74×10^{-2}	0.51	0.94×10^{-2}	0.45
		(0.43×10^{-2})	(0.02)	(0.51×10^{-2})	(0.003)	(0.14×10^{-2})	(0.002)
	5000	0.56×10^{-2}	0.81	0.21×10^{-2}	0.75	0.35×10^{-2}	0.65
		(0.33×10^{-2})	(0.05)	(0.15×10^{-2})	(0.001)	(0.11×10^{-2})	(0.02)

Table 4: Risks and selected cutoff of the penalized procedure with an additional logarithmic term in the penalty. Notation $\mathcal{M}$ stands for the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(4, \frac{1}{2})$.

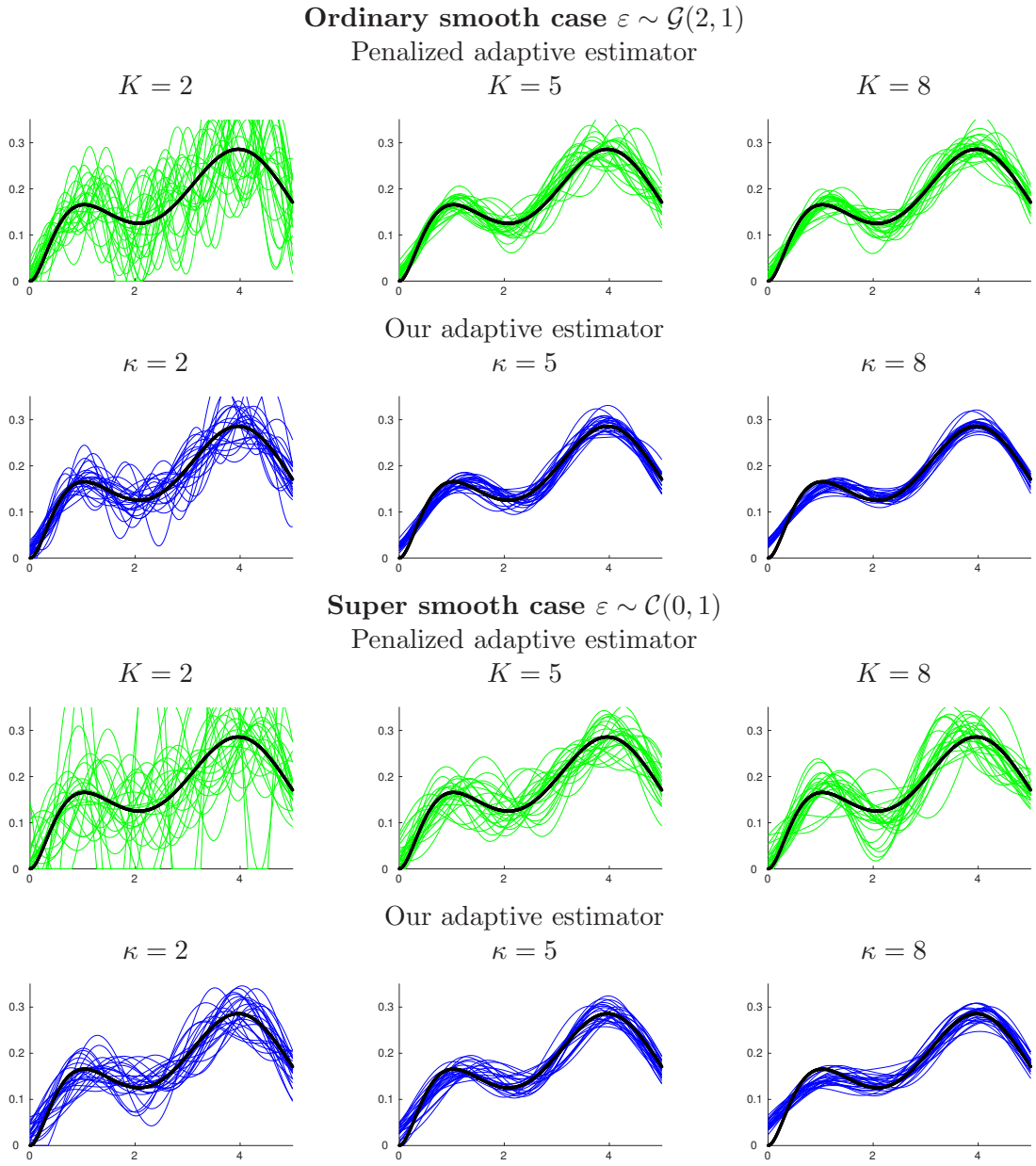

Figure 4: Comparison for different values of K and κ the penalized estimator (green) and our adaptive estimator (blue). Estimation of $f \sim \left(0.3\mathcal{G}(3, \frac{1}{2}) + 0.7\mathcal{G}(4,1)\right)$ (bold black) from $n = 10000$ observations.

3 Decompounding

3.1 Statistical setting

Let Z be a compound Poisson process with intensity $\lambda > 0$ and jump density f, i.e.

$$Z_t := \sum_{j=1}^{N_t} X_j, \quad t \geq 0$$

where N is an homogeneous Poisson process with intensity λ and independent of the i.i.d. variables (X_j) with common density f. One trajectory of Z is observed at sampling rate Δ over $[0, T]$, $T = n\Delta$, $n \in \mathbb{N}$. non-parametric estimation of f, or more generally of the Lvy density λf has been the subject of many papers, among others, [6, 11, 13, 24, 27] for decompounding, [34] for the multidimensional setting, and [3, 14, 33, 37, 48]; a review is also available in the textbook [4] for the non-parametric estimation of the Lvy density.

We observe Z at the time points $j\Delta$, $j = 1, \ldots, n$, for $\Delta > 0$, denote the $j-$th increment by $Y_{j\Delta} = Z_{j\Delta} - Z_{(j-1)\Delta}$. We aim at estimating f from the increments $(Y_{j\Delta}, j = 1, \ldots, n)$. Consider φ the characteristic function of X_1 and φ_Δ the characteristic function of $Z_\Delta = Y_\Delta$. The Lvy-Kintchine formula relates them as follows

$$\varphi_\Delta(u) = \exp\big(\Delta\lambda(\varphi(u) - 1)\big), \quad u \in \mathbb{R}.$$

As Z is a compound Poisson process, $|\varphi_\Delta|$ is bounded from below by $e^{-2\Delta}$, which remains bounded away from 0. Moreover, if $\mathbb{E}[|X_1|] < \infty$ it holds that φ is differentiable and we can then define the distinguished logarithm of φ_Δ (see Lemma 1 in [26])

$$\varphi(u) = 1 + \frac{1}{\lambda\Delta} \operatorname{Log}\big(\varphi_\Delta(u)\big), \quad \text{where } \operatorname{Log}(\varphi_\Delta(u)) = \int_0^u \frac{\varphi'_\Delta(z)}{\varphi_\Delta(z)}\,\mathrm{d}z, \quad u \in \mathbb{R}. \tag{3.1}$$

For simplicity, we assume that the intensity λ is known: $\lambda = 1$. Following (3.1), an estimator of φ is hence given by

$$\widehat{\varphi}_n(u) = 1 + \frac{1}{\Delta} \operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u)), \quad u \in \mathbb{R} \tag{3.2}$$

with

$$\operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u)) := \int_0^u \frac{\widehat{\varphi}'_{\Delta,n}(z)}{\widehat{\varphi}_{\Delta,n}(z)}\,\mathrm{d}z, \quad \widehat{\varphi}_{\Delta,n}(z) = \frac{1}{n}\sum_{j=1}^n e^{izY_{j\Delta}} \text{ and } \widehat{\varphi}'_{\Delta,n}(z) = \frac{1}{n}\sum_{j=1}^n iY_{j\Delta}e^{izY_{j\Delta}}.$$

The quantity $\operatorname{Log}(\widehat{\varphi}_{\Delta,n})$ appearing in $\widehat{\varphi}_n$ might be unbounded: if φ_Δ never cancels, it may not be the case of its estimator $\widehat{\varphi}_{\Delta,n}$. Usually, to prevent this issue a local threshold is used and $\frac{1}{\widehat{\varphi}_{\Delta,n}(v)}$ is replaced with $\frac{1}{\widehat{\varphi}_{\Delta,n}(v)}\mathbb{1}_{|\widehat{\varphi}_{\Delta,n}(v)|>r_n}$, for some vanishing sequence r_n (see e.g. Neumann and Reiß [48]). Here we do not use a local threshold inside the integral, but a global threshold so that $\widehat{\operatorname{Log}(\varphi_{\Delta,n})} = \operatorname{Log}(\widehat{\varphi}_{\Delta,n})$. Define

$$\widetilde{\varphi}_n(u) := \widehat{\varphi}_n(u)\mathbb{1}_{|\widehat{\varphi}_n(u)|\leq 4}, \quad u \in \mathbb{R}, \tag{3.3}$$

where $\widehat{\varphi}_n$ is given by (3.2). The choice of a threshold equal to 4 is technical (see the proof of Theorem 3.1). Cutting off in the spectral domain an applying a Fourier inversion provides the estimator if f

$$\widehat{f}_{m,\Delta}(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\widetilde{\varphi}_n(u)\,\mathrm{d}u, \quad x \in \mathbb{R}. \tag{3.4}$$

3.2 Adaptive upper bound

3.2.1 Upper bound and discussion on the rate.

Theorem 3.1. *Assume that* $\mathbb{E}[X_1^2] < \infty$, $\Delta \leq \frac{1}{4}\log(n\Delta)$ *and* $n\Delta \to \infty$ *as* $n \to \infty$. *Then, for any* $m \geq 0$ *it holds*

$$\mathbb{E}\big[\|\widehat{f}_{m,\Delta} - f\|^2\big] \leq \|f_m - f\|^2 + \frac{2}{n\Delta}\int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + 2\ 5^2\mathbb{E}[X_1^2]\frac{m}{n\Delta} + 2^3 5^2 \frac{m^2}{(n\Delta)^2}.$$

The contraint $\Delta \leq \frac{1}{4}\log(n\Delta)$ is fulfilled for any bounded Δ as $n\Delta \to \infty$. Moreover it allows Δ to be such that $\Delta := \Delta_n \to 0$ and $\Delta_n \to \infty$, not too fast. This last point is interesting. To the knowledge of the author, an estimator that is optimal simultaneously when Δ is fixed or vanishing and consistant, presumably optimal up to a logarithmic loss, when Δ tends to infinity, has not been investigated. Moreover, there are no results on the estimation of the jump density of a compound Poisson process when the sampling rate goes to infinity. In the remaining of this paragraph, we discuss the different rates of convergence implied by Theorem 3.1 according to the behavior of Δ.

Discussion on the rates. The upper bound derived in Theorem 3.1 is the sum of four terms: a bias, plus two variance terms $V \asymp \frac{e^{4\Delta}m}{n\Delta}$ (using that $|\varphi_\Delta(u)| \geq e^{-2\Delta}$) and $V' \asymp \frac{m}{n\Delta}$, which is always smaller or of the same order as V, and a remainder. Assume that f lies is the Sobolev ball $\mathscr{S}(\beta, L)$ (see (2.6)). Then, the bias $\|f - f_m\|^2$ has asymptotic order $m^{-2\beta}$ and we may derive the following rates of convergence.

- **Microscopic and mesoscopic regimes.** Let $\Delta = \Delta_n$ be such that $\Delta_n \to \Delta_0 \in [0, \infty)$ such that $n\Delta_n \to \infty$. Then, the bias variance compromise leads to the choice $m^\star = (e^{-4\Delta_0}n\Delta_0)^{\frac{1}{2\beta+1}}$ and to the rate of convergence $\big(e^{-4\Delta_0}n\Delta_0\big)^{-\frac{2\beta}{2\beta+1}}$ that matches the optimal rates of convergence as Δ_0 is fixed or tending to 0. Indeed, the rate is in $T^{-\frac{2\beta}{2\beta+1}}$, with $T = n\Delta_n$ denoting the time horizon, it is clearly rate optimal as it corresponds to the optimal rate of convergence to estimate the jump density of a compound Poisson process from continuous observations ($\Delta = 0$). The constant $e^{-4\Delta_0}$ appearing in the rate depends exponentially on Δ_0, which asymptotically as little effect but in practice deteriorates the numerical performances.

- **Macroscopic regime.** Let $\Delta = \Delta_n \to \infty$ such that $\Delta_n \leq \frac{1}{4}\log(n\Delta_n)$ The variance term V tends to 0, so that the estimator is consistent. Heuristically, if Δ goes to infinity the central limit theorem states that Y_Δ is close in law to a parametric Gaussian variable,

e.g. if f is centered and with unit variance it holds that: $\sqrt{\Delta}^{-1} Y_\Delta \xrightarrow[\Delta\to\infty]{d} \mathcal{N}(0,1)$. Consequently, the fact that f can be constantly estimated is non trivial. Duval [25] establishes that if $\Delta = O((n\Delta)^\delta)$, for some $\delta \in (0,1)$, i.e. when Δ_n goes rapidly to infinity, there exists no consistent non-parametric estimator of f. The fact that estimation is impossible when Δ goes too rapidly to infinity was established through an asymptotic equivalence result. In this case it is always possible to build two different compound Poisson processes for which the statistical experiments generated by their increments are asymptotically equivalent. Therefore, the result of Theorem 3.1 is new in that context. We may distinguish two additional regimes:

1. Slow macroscopic regime. If $\Delta_n = o\big(\log(n\Delta_n)\big)$, the choice $m^\star = \big(e^{-4\Delta_n} n\Delta_n\big)^{\frac{1}{2\beta+1}}$ leads to the rate of convergence $\big(e^{-4\Delta_n} n\Delta_n\big)^{-\frac{2\beta}{2\beta+1}}$. There is no lower bound in the literature to ensure if this rate is optimal. However if Δ goes slowly to infinity, for example if $\Delta_n = \log(\log(n\Delta_n))$, then the rate is $\big((\log(n\Delta_n))^{-4} n\Delta_n\big)^{-\frac{2\beta}{2\beta+1}}$, which is rate optimal, up to the logarithmic loss that may not be optimal.
2. Intermediate macroscopic regime. Let $\Delta_n = \delta \log(n\Delta_n)$, $0 < \delta < 1/4$, then $m^\star = (n\Delta_n)^{\frac{1-4\delta}{2\beta+1}}$, leading to the rate $(n\Delta_n)^{-\frac{2\beta(1-4\delta)}{2\beta+1}}$. This rate deteriorates as δ increases. The limit $\delta = 1/4$ imposed by Theorem 3.1 may not be optimal, no lower bound adapted to this case exists in the literature.

The interest of the macroscopic regime is mainly theoretical as in practice if Δ is a large constant to get $e^{-4\Delta} n\Delta$ large one should consider a huge amount n of observations. However, this regime enlightens the role of the sampling rate Δ in the non-parametric estimation of the jump density.

3.2.2 Adaptive choice of the cutoff parameter

We consider the optimal cutoff $\overline{m}_n$ given by

$$\overline{m}_n \in \operatorname*{arginf}_{m\geq 0}\Big\{\|f_m - f\|^2 + \frac{2}{n\Delta}\int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + 2\ 5^2 \mathbb{E}[X_1^2]\frac{m}{n\Delta} + 2^3 5^2 \frac{m^2}{(n\Delta)^2}\Big\}.$$

Following the previous strategy, the upper bound given by Theorem 3.1 is optimal, at least for $\Delta \to [0,\infty)$. The leading variance terms is in $\frac{me^{4\Delta}}{n\Delta}$, we differentiate in m the upper bound to find that the optimal cutoff $m^\star \asymp \overline{m}_n$ is such that:

$$|\varphi(\overline{m}_n)|^2 = \frac{e^{4\Delta}}{\Delta n} \Leftrightarrow |\varphi(\overline{m}_n)|^2 = \frac{e^{4\Delta}}{\Delta n},$$

which has an empirical version, we select $\widehat{m}_n$ accordingly. As in the deconvolution setting, we modify the estimator $\widetilde{\varphi}_n$ in (3.3) which is set to 0 when the estimator of $|\varphi|$ is smaller that $1/\sqrt{n\Delta}$, meaning that the noise is dominant. Define

$$\overline{\varphi}_n(u) := \widetilde{\varphi}_n(u) \mathbb{1}_{|\widetilde{\varphi}_n(u)| \geq \kappa_{n,\Delta}/\sqrt{n\Delta}}, \quad u \in \mathbb{R}$$

where $\kappa_{n,\Delta} := (e^{2\Delta} + \kappa\sqrt{\log(n\Delta)})$, $\kappa > 0$, and the new the estimator if f

$$\overline{f}_{m,\Delta}(x) = \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\overline{\varphi}_n(u)\,\mathrm{d}u, \quad x \in \mathbb{R}.$$

Again, Lemma 2.1 ensures that Theorem 3.1 holds for the estimator $\overline{f}_{m,\Delta}$. Finally, we introduce the empirical threshold, for some $\alpha \in (0,1]$ and $\kappa > 0$

$$\widehat{m}_n = \max\Big\{m \geq 0 : |\overline{\varphi}_n(m)| = \frac{\kappa_{n,\Delta}}{\sqrt{n\Delta}}\Big\} \wedge (n\Delta)^{\alpha}.$$

Theorem 3.2. *Assume that $\mathbb{E}[X_1^4] < \infty$, $\Delta \leq \frac{1}{4}\log(n\Delta)$ and $n\Delta \to \infty$ as $n \to \infty$. Then, for a positive constant C_1, depending on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$, and C_2 a constant depending on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$, it holds*

$$\mathbb{E}[\|\overline{f}_{\widehat{m}_n,\Delta} - f\|^2] \leq C_1 \inf_{m\in[0,(n\Delta)^{\alpha}]}\Big\{\|f_m - f\|^2 + \frac{\log(n\Delta)m}{n\Delta} + \frac{1}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_{\Delta}(u)|^2} + \frac{m^2}{(n\Delta)^2}\Big\}$$
$$+ C_2\Big(\frac{1}{n\Delta} + (n\Delta)^{\alpha-\kappa^2\Delta^2 e^{-4\Delta}}\Big).$$

If $\kappa > \frac{\sqrt{2}e^{2\Delta}}{\Delta}$ the last additional term is negligible, regardless the value $\alpha \leq 1$ and Theorem 3.2 ensures that the adaptive estimator $\overline{f}_{\widehat{m}_n,\Delta}$ satisfies the same upper bound as in Theorem 3.1. Therefore, it is adaptive and rate optimal, up to a logarithmic term and the multiplicative constant C_1, in the microscopic and mesoscopic regimes defined above. In the macroscopic regimes such that $\Delta := \Delta_n \to \infty$ such that $\Delta_n < \frac{1}{4}\log(n\Delta_n)$ as $n \to \infty$ the estimator is consistent. Note that to establish the adaptive upper bound we imposed a stronger assumption that $\mathbb{E}[X_1^4] < \infty$. In the following numerical study, we recover that the procedure is stable in the choice of κ.

3.3 Numerical results

As for the deconvolution problem, we illustrate the performance of this adaptive estimator for different densities f. We consider the same densities as for the deconvolution problem, the Cauchy density excepted as it is not covered by our procedure: it has infinite moments. We compute the adaptive L^2-risks of our procedure over 1000 Monte Carlo iterations for various values of κ. We consider $n = 5000$ and the sampling interval $\Delta = 1$. The results are represented on Figure 5, we observe that the rates are small and stable regardless the value of κ and the density considered.

4 Concluding remarks

Comments on the adaptive procedure. In the present paper we develop an adaptive procedure that was successfully used in of Duval and Kappus [26] which considers the problem of grouped data estimation. One observes i.i.d. realizations of $Y_j = X_j^{(1)} + \ldots + X_j^{(K)}$ where

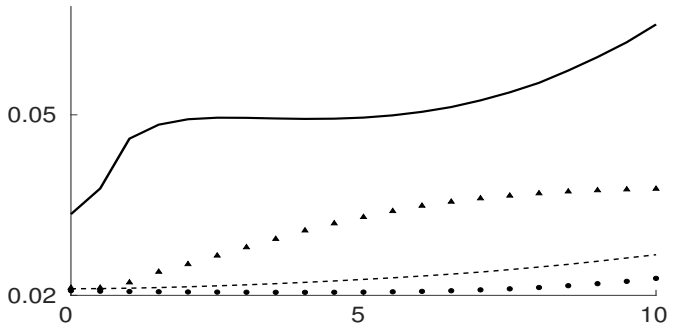

Figure 5: **Decompounding:** Computations by $M = 1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa \leq \frac{1}{2}(\sqrt{n}-1)\sqrt{\log(n)}^{-1} = 11.9$ (x axis). Estimation of f from $n = 5000$ ($T = 5000$ and $\Delta = 1$) increments of a compound Poisson process with intensity $\lambda = 1$ and jump density f: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2,\frac{1}{2})$ (triangles).

K is a fixed and known integer and where the random variables $(X_j^{(1)}, \ldots, X_j^{(K)})_{1\leq j\leq n}$ are i.i.d. This problem is a particular deconvolution problem where the density of the noise is unknown and depends on the density of interest. Here we adapt the procedure to two other classical inverse problems: deconvolution and decompounding. In each cases the resulting adaptive estimator is proven rate optimal, up to a logarithmic factor, for the L^2-risk.

Both in the grouped data setting (see [26]) and in the deconvolution setting (see Section 2) the computation of the adaptive cutoff, after simplifications, involves the set

$$\widehat{m}_n \in \left\{ |\widehat{\varphi}_{Y,n}(u)| = \frac{1}{\sqrt{n}}(1 + \kappa\sqrt{\log n}) \right\}, \quad \kappa > 0 \tag{4.1}$$

and not the characteristic function of X_1, nor the one of the errors in the deconvolution problem, nor the inverse of the operator relation φ_Y to φ_X. However there are some differences between the grouped data setting where a uniform control on the characteristic function was needed and the deconvolution framework. We only have pointwise control here, this difference is due to the fact that in the grouped data setting the density f both played the role of the quantity of interest and the density of the noise; in [26] we needed to bound it from above and below.

In the decoumpounding setting the computation of the adaptive cutoff involves the empirical characteristic function of the jump density, which is more challenging than in the latter case. It does not seem obvious to have a simplified version as (4.1) even though numerically selecting $\widehat{m}_n$ this way seems relevant.

Decompounding setting. In the present paper we exhibit an adaptive rate optimal (up to a logarithmic factor) estimator of the jump density of a compound Poisson process from the discrete observation of its increments at sampling rate $\Delta = \Delta_n \to [0, \infty)$. It allows $\Delta_n \to \infty$ such that $\Delta_n < 1/4 \log(n\Delta_n)$, our estimator remains consistent and optimal up

to a logarithmic factor in some cases, e.g. if $\Delta_n = O\big(\log\log(n\Delta_n)\big)$. Using [25], consistent non-parametric estimation of the jump density is impossible if $\exists \delta > 0,\ \Delta_n = O(n\Delta_n)^{\delta}$, the remaining questions are what happens in between and if the log loss in the upper bound that appears when $\Delta_n \to \infty$ is avoidable or not. The constant $1/4$ in the constant $\Delta_n < 1/4\log(n\Delta_n)$ of Theorem 3.1 can probably be improved.

5 Proofs

5.1 Proof of Theorem 2.1

Let $m \geq 0$ be fixed. First, consider the event $\mathcal{E} = \{\widehat{m}_n < m\}$, on this event we control the surplus in the bias of the estimator $\widetilde{f}_{\widehat{m}_n}$. Using the inequality

$$|\varphi_X|^2 \leq 2\frac{|\widehat{\varphi}_{Y,n}|^2}{|\varphi_\varepsilon|^2} + 2\frac{|\varphi_Y - \widehat{\varphi}_{Y,n}|^2}{|\varphi_\varepsilon|^2},$$

along with the definition of $\widehat{m}_n$, gives

$$\begin{aligned}\mathbb{E}\Big[\mathbb{1}_{\mathcal{E}} \int\limits_{|u|\in[\widehat{m}_n,m]} |\varphi_X(u)|^2\,\mathrm{d}u\Big] &\leq 2\mathbb{E}\Big[\mathbb{1}_{\mathcal{E}} \int\limits_{|u|\in[\widehat{m}_n,m]} \frac{\kappa_n^2 n^{-1}}{|\varphi_\varepsilon(u)|^2}\,\mathrm{d}u\Big] + \int_{-m}^{m} \frac{\mathbb{E}[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2}\,\mathrm{d}u \\ &\leq \int_{-m}^{m} \frac{2(\kappa_n^2+1)n^{-1}}{|\varphi_\varepsilon(u)|^2}\,\mathrm{d}u.\end{aligned}$$

Recall that $\kappa_n = 1 + \kappa\sqrt{\log(n)}$ and (2.3). This implies immediately that, on the event $\mathcal{E}$, for a positive constant C depending only on the choice of κ,

$$\mathbb{E}[\|\widetilde{f}_{\widehat{m}_n} - f\|^2 \mathbb{1}_{\mathcal{E}}] \leq \|f_m - f\|^2 + C\frac{\log n}{n}\int_{-m}^{m} \frac{\mathrm{d}u.}{|\varphi_\varepsilon(u)|^2}.$$

Second, consider the complement set $\mathcal{E}^c$, where we control the surplus in the variance of $\widetilde{f}_{\widehat{m}_n}$. By the definition of $\widehat{m}_n$, it holds

$$\begin{aligned}&\mathbb{E}\Big[\int\limits_{|u|\in[m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbb{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}}\,\mathrm{d}u \mathbb{1}_{\mathcal{E}^c}\Big] \\ &\qquad\leq \int\limits_{|u|\in[m,n^\alpha]} \frac{\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2} \mathbb{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}}\,\mathrm{d}u.\end{aligned}$$

On the event $\{|\varphi_Y(u)| > n^{-1/2}\}$, we derive that

$$\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2] \leq |\varphi_Y(u)|^2 + \mathbb{E}[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2] \leq |\varphi_Y(u)|^2 + \frac{1}{n} \leq 2|\varphi_Y(u)|^2.$$

Consequently, we get

$$\mathbb{E}\Big[\int\limits_{|u|\in[m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2\,\mathbb{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}}\,\mathrm{d}u\,\mathbb{1}_{\mathcal{E}^c}\Big] \leq 2\int_{[-m,m]^c} |\varphi_X(u)|^2\,\mathrm{d}u.$$

Next, using that $|\widetilde{\varphi}_{X,n}(u)| \le 1$, we derive that

$$\begin{aligned}
&\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbb{1}_{\{|\varphi_Y(u)|\le n^{-1/2}\}} \,\mathrm{d}u \, \mathbb{1}_{\mathcal{E}^c} \Big] \\
&\qquad \le \int_{|u|\in[m,n^\alpha]} |\varphi_X(u)|^2 \,\mathrm{d}u + 4 \int_{|u|\in[m,n^\alpha]} \mathbb{P}(|\widehat{\varphi}_{Y,n}(u)| \ge \kappa_n n^{-1/2}) \mathbb{1}_{\{|\varphi_Y(u)|\le n^{-1/2}\}} \,\mathrm{d}u \\
&\qquad \le \int_{|u|\in[m,n^\alpha]} |\varphi_X(u)|^2 \,\mathrm{d}u + 4 \int_{|u|\in[m,n^\alpha]} \mathbb{P}(|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)| > \kappa(\log n/n)^{1/2}) \,\mathrm{d}u \\
&\qquad \le \int_{u\in[m,m]^c} |\varphi_X(u)|^2 \,\mathrm{d}u + 8 n^{\alpha-\kappa^2/2}.
\end{aligned}$$

The last inequality is a direct consequence of the Hoeffding inequality. Putting the above together, we have shown that for universal positive constants C_1 and C_3 and a constant C_2 depending only on κ, for all $m \ge 0$,

$$\mathbb{E}[\|\widetilde{f}_{\widehat{m}_n} - f\|^2] \le C_1 \|f - f_m\|^2 + C_2 \frac{\log n}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} + C_3 n^{\alpha-\kappa^2/2}.$$

Taking the infimum over m completes the proof. □

5.2 Proof of Theorem 3.1

Proof of Theorem 3.1 uses similar arguments as the proof of Theorem 1 of Duval and Kappus [26]. However, estimator (3.4) is different from the estimator studied in [26] and we need to take into account the additional parameter Δ that needs to be carefully handled.

5.2.1 Preliminaries

We establish two technical Lemmas used in the proof of Theorem 3.1.

Lemma 5.1. *Let $m > 0$ and $\zeta > 0$ and define the event*

$$\Omega_{\zeta,\Delta}(m) := \Big\{ \forall u \in [-m,m],\ \big|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)\big| \le \zeta \sqrt{\frac{\log(n\Delta)}{n\Delta}} \Big\}.$$

1. *If $\mathbb{E}[X_1^2]$ is finite, then, the following holds for $\eta > 0$ and any $\zeta > \sqrt{\Delta(1+2\eta)}$,*

$$\mathbb{P}\big(\Omega_{\zeta,\Delta}(m)^c\big) \le \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4\frac{m}{(n\Delta)^\eta}.$$

2. *If $\mathbb{E}[X_1^4]$ is finite, then, the following holds for $\eta > 0$ and any $\zeta > \sqrt{\Delta(1+2\eta)}$,*

$$\mathbb{P}\big(\Omega_{\zeta,\Delta}(m)^c\big) \le \frac{C}{(n\Delta)^2} + 4\frac{m}{(n\Delta)^\eta},$$

where C depends on $\mathbb{E}[X_1^4]$ and $\mathbb{E}[X_1^2]$.

Proof of Lemma 5.1. Consider the events

$$A(c) := \Big\{ \Big| \frac{1}{n} \sum_{j=1}^{n} |Y_{j\Delta}| - \mathbb{E}[|Y_\Delta|] \Big| \le c \Big\}$$

$$B_{h,\tau}(m) := \Big\{ \forall |k| \le \Big\lceil \frac{m}{h} \Big\rceil,\ \big| \widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh) \big| \le \tau \sqrt{\frac{\log(n\Delta)}{n\Delta}} \Big\}$$

for some positive constants c, h and τ to be determined. First, using that $x \to e^{iux}$ is 1-Lipschitz and that $\mathbb{E}[|Y_\Delta|] \le \Delta \mathbb{E}[|X_1|]$ we get on the event $A(c)$

$$\big| \widehat{\varphi}_{\Delta,n}(u) - \widehat{\varphi}_{\Delta,n}(u+h) \big| \mathbb{1}_{A(c)} \le h\big(\Delta \mathbb{E}[|X_1|] + c \big), \quad \forall u \in \mathbb{R},\ h > 0. \tag{5.1}$$

If $\mathbb{E}[X_1^2]$ is finite the Markov inequality and the bound $\mathbb{V}[|Y_\Delta|] \le \mathbb{V}[Y_\Delta] = \Delta \mathbb{E}[X_1^2]$ lead to

$$\mathbb{P}\big(A(c)^c \big) \le \frac{\Delta \mathbb{E}[X_1^2]}{c^2 n}. \tag{5.2}$$

If $\mathbb{E}[X_1^4]$ is finite (5.2) can be improved using that

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{n} (|Y_{j\Delta}| - \mathbb{E}[|Y_\Delta|]) \Big)^4 \Big] \le n\Delta^2 \mathbb{E}[X_1^4] + 3n(n-1)\Delta^2 \mathbb{E}[X_1^2]$$

, leading to

$$\mathbb{P}\big(A(c)^c \big) \le C \frac{\Delta^2}{c^4 n^2}, \tag{5.3}$$

where C is a constant depending on $\mathbb{E}[X_1^4]$ and $\mathbb{E}[X_1^2]$. Second, we have that

$$\begin{aligned}
\mathbb{P}\big(B_{h,\tau}(m)^c \big) &\le \mathbb{P}\Big(\exists\ |k| \le \Big\lceil \frac{m}{h} \Big\rceil,\ \big| \widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh) \big| > \tau \sqrt{\frac{\log(n\Delta)}{n\Delta}} \Big) \\
&\le \sum_{k=-\lceil m/h \rceil}^{\lceil m/h \rceil} \mathbb{P}\Big(\big| \widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh) \big| > \tau \sqrt{\frac{\log(n\Delta)}{n\Delta}} \Big) \\
&\le \sum_{k=-\lceil m/h \rceil}^{\lceil m/h \rceil} 2 \exp\Big(- \frac{\tau^2 \log(n\Delta)}{2\Delta} \Big) = 4 \Big\lceil \frac{m}{h} \Big\rceil (n\Delta)^{-\tau^2/(2\Delta)}
\end{aligned}$$

where the last inequality is obtained applying the Hoeffding inequality. Let $|u| \le m$, there exists k such that $u \in [kh - \frac{h}{2}, kh + \frac{h}{2}]$ and we can write that

$$\begin{aligned}
\mathbb{1}_{A(c) \cap B_{h,\tau}(m)} \big| \widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u) \big| \le \mathbb{1}_{A(c) \cap B_{h,\tau}(m)} \Big(& \big| \widehat{\varphi}_{\Delta,n}(u) - \widehat{\varphi}_{\Delta,n}(kh) \big| + \big| \widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh) \big| \\
& + \big| \varphi_\Delta(kh) - \varphi_\Delta(u) \big| \Big).
\end{aligned}$$

Using (5.1), the definition of $B_{h,\tau}(m)$ and that $x \to e^{iux}$ is 1-Lipschitz, lead to

$$\mathbb{1}_{A(c)\cap B_{h,\tau}(m)} \sup_{u\in[-m,m]} \left|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)\right| \le 2h\Delta\mathbb{E}[|X_1|] + hc + \tau\sqrt{\frac{\log(n\Delta)}{n\Delta}}. \tag{5.4}$$

Taking $c = \Delta$, $h = o\big(\sqrt{\frac{\log(n\Delta)}{n\Delta}}\big)$ such that $h > 1/\sqrt{n\Delta}$ and $\zeta > \tau$, (5.4) shows that $A(c) \cap B_{h,\tau}(m) \subset \Omega_{\zeta,\Delta}(m)$. Moreover, it follows from $h > 1/\sqrt{n\Delta}$, (5.1) and (5.2) that, for all $\eta > 0$

$$\mathbb{P}\big(\Omega^c_{\zeta,\Delta}(m)\big) \le \mathbb{P}\big(A^c(\Delta)\big) + \mathbb{P}\big(B^c_{h,\tau}(m)\big) \le \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4\left\lceil\frac{m}{h}\right\rceil(n\Delta)^{-\frac{\tau^2}{2\Delta}} \le \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4m(n\Delta)^{\frac{\Delta-\tau^2}{2\Delta}}.$$

Finally, choosing $\tau^2 = \Delta(1+2\eta)$ leads to the result. The second inequality is obtained follows from similar arguments using (5.2) instead of (5.3). □

Lemma 5.2. *Let $\gamma > 0$, define*

$$M^{(\gamma)}_{n,\Delta} := \min\big\{m \ge 0 : |\varphi_\Delta(m)| = \gamma\sqrt{\log(n\Delta)/(n\Delta)}\big\},$$

with the convention $\inf\{\emptyset\} = +\infty$. *Take $\gamma > \zeta > 0$, then, we have*

$$\mathbb{1}_{|u|\le M^\gamma_{n,\Delta}\wedge m,\Omega_{\zeta,\Delta}(m)}\Big|\operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \operatorname{Log}(\varphi_\Delta(u))\Big| \le \frac{\gamma}{\zeta}\log\Big(\frac{\gamma}{\gamma-\zeta}\Big)\frac{|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|}{|\varphi_\Delta(u)|}.$$

Proof of Lemma 5.2. First note that for $|u| \le M^\gamma_{n,\Delta}$, the ratio $\frac{\varphi'_\Delta}{\varphi_\Delta}$ is well defined. Moreover, on the event $\Omega_{\zeta,\Delta}(m)$ then we have that

$$|\widehat{\varphi}_{\Delta,n}(u)| \ge |\varphi_\Delta(u)| - |\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \ge (\gamma-\zeta)\sqrt{\frac{\log(n\Delta)}{n\Delta}} > 0, \quad \forall |u| \le m.$$

Then, the quantity $\frac{\widehat{\varphi}'_{\Delta,n}}{\widehat{\varphi}_{\Delta,n}}$ is also well defined if $\gamma > \zeta$. For $v \in \mathbb{R}$, notice that

$$\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)} = \frac{\Big(-\frac{(\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v))}{\varphi_\Delta(v)}\Big)'}{\Big(1 - \frac{(\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v))}{\varphi_\Delta(v)}\Big)}. \tag{5.5}$$

On the event $\Omega_{\zeta,\Delta}(m)$, it holds $\forall u \in [-m\wedge M^\gamma_{n,\Delta}, m\wedge M^\gamma_{n,\Delta}]$

$$|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \le \zeta\sqrt{\log(n\Delta)}(n\Delta)^{-\frac{1}{2}} \le \frac{\zeta}{\gamma}|\varphi_\Delta(u)|, \tag{5.6}$$

where $\gamma > \zeta$. Then, a Neumann series expansion, with (5.5) and (5.6) gives for $|v| \le m\wedge M^\gamma_{n,\Delta}$,

$$\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)} = -\sum_{\ell=0}^{\infty}\Big(\frac{\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v)}{\varphi_\Delta(v)}\Big)'\Big(\frac{\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v)}{\varphi_\Delta(v)}\Big)^\ell,$$

where

$$\Big(\frac{\widehat{\varphi}(v)-\varphi_\Delta(v)}{\varphi_\Delta(v)}\Big)'\Big(\frac{\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v)}{\varphi_\Delta(v)}\Big)^\ell=\frac{1}{\ell+1}\Big[\Big(\frac{\widehat{\varphi}_{\Delta,n}(v)-\varphi_\Delta(v)}{\varphi_\Delta(v)}\Big)^{\ell+1}\Big]'.$$

Using $\widehat{\varphi}_\Delta(0)-\varphi_\Delta(0)=0$ and (5.6), we get

$$\begin{aligned}\mathbb{1}_{|u|\le m\wedge M_{n,\Delta}^\gamma,\Omega_{\zeta,\Delta}(m)}\Big|\int_0^u\Big(\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)}-\frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)}\Big)\,\mathrm{d}v\Big|&\le\sum_{\ell=0}^\infty\frac{1}{\ell+1}\frac{|\widehat{\varphi}_{\Delta,n}(u)-\varphi_\Delta(u)|^{\ell+1}}{|\varphi_\Delta(u)|^{\ell+1}}\\&\le\frac{|\widehat{\varphi}_{\Delta,n}(u)-\varphi_\Delta(u)|}{|\varphi_\Delta(u)|}\sum_{\ell=0}^\infty\frac{(\zeta/\gamma)^\ell}{\ell+1}\\&=\frac{\gamma}{\zeta}\log\big(\frac{\gamma}{\gamma-\zeta}\big)\frac{|\widehat{\varphi}_{\Delta,n}(u)-\varphi_\Delta(u)|}{|\varphi_\Delta(u)|},\end{aligned}\tag{5.7}$$

which completes the proof. □

5.2.2 Proof of Theorem 3.1

We have the decomposition

$$\|\widehat{f}_{m,\Delta}-f\|^2=\|f_m-f\|^2+\|\widehat{f}_{m,\Delta}-f_m\|^2=\|f_m-f\|^2+\frac{1}{2\pi}\int_{-m}^m|\widetilde{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u.$$

Let $\gamma>\zeta$, we decompose the second term on the events $\{m\le M_{n,\Delta}^\gamma\}$ and $\Omega_{\zeta,\Delta}(m)$ of Lemma 5.2,

$$\begin{aligned}\int_{-m}^m|\widetilde{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u&=\int_{-m\wedge M_{n,\Delta}^\gamma}^{m\wedge M_{n,\Delta}^\gamma}\mathbb{1}_{\Omega_{\zeta,\Delta}(m)}|\widetilde{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u\\&\quad+\mathbb{1}_{m>M_{n,\Delta}^\gamma,\Omega_{\zeta,\Delta}(m)}\int_{|u|\in[M_{n,\Delta}^\gamma,m]}|\widetilde{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u\\&\quad+\mathbb{1}_{\Omega_{\zeta,\Delta}(m)^c}\int_{-m}^m|\widetilde{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u\\&:=T_{1,n}+T_{2,n}+T_{3,n}.\end{aligned}$$

Fix $\gamma_\Delta=\frac{2\zeta}{1\wedge\Delta}>\zeta$. On the event $\{|u|\le m\wedge M_{n,\Delta}^{\gamma_\Delta},\Omega_{\zeta,\Delta}(m)\}$, Lemma 5.2 and equations (3.1), (3.2) and (5.7), along with (5.6), imply

$$\big|\widehat{\varphi}_n(u)\big|\le1+\frac{|\operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u))-\operatorname{Log}(\varphi_\Delta(u))|+|\operatorname{Log}(\varphi_\Delta(u))|}{\Delta}\le3+\frac{1}{\Delta}\log\Big(\frac{\gamma}{\gamma-\zeta}\Big)\le4,$$

consequently $\widetilde{\varphi}_n(u)=\widehat{\varphi}_n(u)$. Then, we get from Lemma 5.2 and the definition of γ_Δ, that

$$\begin{aligned}\mathbb{E}[T_{1,n}]&=\frac{1}{\Delta^2}\int_{-m\wedge M_{n,\Delta}^{\gamma_\Delta}}^{m\wedge M_{n,\Delta}^{\gamma_\Delta}}\mathbb{E}\Big[\mathbb{1}_{\Omega_{\zeta,\Delta}(m)}\big|\operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u))-\operatorname{Log}(\varphi_\Delta(u))\big|^2\Big]\,\mathrm{d}u\\&\le\frac{1}{\Delta^2}\int_{-m}^m\frac{\mathbb{E}\big[|\widehat{\varphi}_{\Delta,n}(u)-\varphi_\Delta(u)|^2\big]}{|\varphi_\Delta(u)|^2}\,\mathrm{d}u.\end{aligned}$$

Direct computations together with the Lvy-Kintchine formula lead to

$$\mathbb{E}\big[|\widehat{\varphi}_{\Delta,n}(u)-\varphi_\Delta(u)|^2\big]=\frac{1-|\varphi_\Delta(u)|^2}{n}\le\frac{\big(2\Delta|\mathrm{Re}(\varphi(u))-1|\big)\wedge 1}{n}\le\frac{2\Delta\wedge 1}{n}.$$

We derive that

$$\mathbb{E}[T_{1,n}]\le\frac{2\Delta\wedge 1}{n\Delta^2}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}\le\frac{2}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}.$$

Next, fix $\zeta>\sqrt{5\Delta}$, Lemma 5.1 with $\eta=2$ gives

$$\mathbb{E}[T_{3,n}]\le 2\ 5^2 m\Big(\frac{\mathbb{E}[X_1^2]}{n\Delta}+4\frac{m}{(n\Delta)^2}\Big).$$

Moreover, using $|\varphi_\Delta(u)|\ge e^{-2\Delta}$, $\forall u\in\mathbb{R}$ together with the constraint $\Delta\le\delta\log(n\Delta)$, $\delta<\frac{1}{4}$, we get

$$|\varphi_\Delta(u)|\ge(n\Delta)^{-2\delta}>\gamma_\Delta\sqrt{\log(n\Delta)/(n\Delta)},\quad\forall u\in\mathbb{R}.$$

Finally, $M_{n,\Delta}^{\gamma_\Delta}=+\infty$, $\forall\zeta>0$ and $T_{2,n}=0$ almost surely. Gathering all terms completes the proof. □

5.3 Proof of Theorem 3.2

Let $m\ge 0$ be fixed. Consider the event $\mathcal{E}=\{\widehat{m}_n<m\}$, on this event we control the surplus in the bias of the estimator $\widetilde{f}_{\widehat{m}_n}$. Using the inequality $|\varphi|^2\le 2|\widetilde{\varphi}_n|^2+2|\varphi-\widetilde{\varphi}_n|^2$, along with the definition of $\widehat{m}_n$ and Theorem 3.1, gives

$$\begin{aligned}\mathbb{E}\Big[\mathbb{1}_{\mathcal{E}}\int_{|u|\in[\widehat{m}_n,m]}|\varphi(u)|^2\,\mathrm{d}u\Big]&\le 2\mathbb{E}\Big[\mathbb{1}_{\mathcal{E}}\int_{|u|\in[\widehat{m}_n,m]}\frac{\kappa_{n,\Delta}^2}{n\Delta}\,\mathrm{d}u\Big]+2\int_{-m}^{m}\mathbb{E}[|\widetilde{\varphi}_n(u)-\varphi(u)|^2]\,\mathrm{d}u\\&\le 4\frac{\kappa_{n,\Delta}^2 m}{n\Delta}+\frac{4}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}+2^2 5^2\mathbb{E}[X_1^2]\frac{m}{n\Delta}+2^4 5^2\frac{m^2}{(n\Delta)^2}.\end{aligned}$$

Recall that $\kappa_{n,\Delta}=e^{2\Delta}+\kappa\sqrt{\log(n\Delta)}$ together with Theorem 3.1, this implies immediately that, on the event $\mathcal{E}$, for a positive constant C depending on the choice of κ and $\mathbb{E}[X_1^2]$,

$$\mathbb{E}[\|\overline{f}_{\widehat{m}_n,\Delta}-f\|^2\mathbb{1}_{\mathcal{E}}]\le\|f_m-f\|^2+C\frac{\log(n\Delta)m}{n\Delta}+\frac{5}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}+2^5 5^2\frac{m^2}{(n\Delta)^2}.$$

Second, consider the complement set $\mathcal{E}^c$, where we control the surplus in the variance of $\widetilde{f}_{\widehat{m}_n}$. By the definition of $\widehat{m}_n$, it holds

$$\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]}|\overline{\varphi}_n(u)-\varphi(u)|^2\,\mathrm{d}u\mathbb{1}_{\mathcal{E}^c}\Big]\le\int_{|u|\in[m,(n\Delta)^\alpha]}\mathbb{E}[|\overline{\varphi}_n(u)-\varphi(u)|^2]\,\mathrm{d}u.$$

Let $\eta>2$, such that $\alpha-\eta<-1$, and $\zeta>\sqrt{\Delta(1+2\eta)}$ and γ_Δ as in the proof of Theorem 3.1 (leading to $M_{n,\Delta}^{\gamma_\Delta}=+\infty$). Then, Lemmas 5.1 (decomposing on $\Omega_{\zeta,\Delta}((n\Delta)^\alpha)$) and 5.2 lead to

$$\mathbb{E}[|\overline{\varphi}_n(u)-\varphi(u)|^2]\le|\varphi(u)|^2+\mathbb{E}[|\widehat{\varphi}_n(u)-\varphi(u)|^2]\le|\varphi(u)|^2+\frac{2}{n\Delta|\varphi_\Delta(u)|^2}+\frac{\mathbb{E}[X_1^2]}{n\Delta}+4\frac{1}{n\Delta}.$$

First, on the event $\{|\varphi(u)| > e^{2\Delta}/\sqrt{n\Delta}\}$, we obtain

$$\mathbb{E}[|\overline{\varphi}_n(u) - \varphi(u)|^2] \le |\varphi(u)|^2\big(6 + \mathbb{E}[X_1^2]\big).$$

Consequently, define $C_0 := 6 + \mathbb{E}[X_1^2]$, then,

$$\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\overline{\varphi}_n(u) - \varphi(u)|^2 \mathbb{1}_{\{|\varphi(u)|>e^{2\Delta}/\sqrt{n\Delta}\}}\,\mathrm{d}u\,\mathbb{1}_{\mathcal{E}^c}\Big] \le C_0 \int_{[-m,m]^c} |\varphi(u)|^2\,\mathrm{d}u.$$

Next, using that $|\overline{\varphi}_n(u)| \le 4$ and the definition of $\widehat{m}_n$, we derive that

$$\begin{aligned}
&\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\overline{\varphi}_n(u) - \varphi(u)|^2 \mathbb{1}_{\{|\varphi(u)|\le e^{2\Delta}/\sqrt{n\Delta}\}}\,\mathrm{d}u\,\mathbb{1}_{\mathcal{E}^c}\Big] \\
&\qquad \le \int_{|u|\in[m,(n\Delta)^\alpha]} |\varphi(u)|^2\,\mathrm{d}u + 5^2 \int_{|u|\in[m,(n\Delta)^\alpha]} \mathbb{P}\big(|\widehat{\varphi}_n(u)| \ge \kappa_{n,\Delta}/\sqrt{n\Delta}\big)\mathbb{1}_{\{|\varphi(u)|\le e^{2\Delta}/\sqrt{n\Delta}\}}\,\mathrm{d}u \\
&\qquad \le \int_{|u|\in[m,(n\Delta)^\alpha]} |\varphi(u)|^2\,\mathrm{d}u + 5^2 \int_{|u|\in[m,(n\Delta)^\alpha]} \mathbb{P}\big(|\widehat{\varphi}_n(u) - \varphi(u)| > \kappa\sqrt{\log(n\Delta)/(n\Delta)}\big)\,\mathrm{d}u \\
&\qquad \le \int_{u\in[m,m]^c} |\varphi(u)|^2\,\mathrm{d}u + 5^2 T_n.
\end{aligned}$$

Finally, we give a bound for T_n using Lemmas 5.1 and 5.2 with γ_Δ and $\zeta > 0$ as above,

$$\begin{aligned}
\mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \ge \kappa\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) &= \mathbb{P}\Big(|\operatorname{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \operatorname{Log}(\varphi_\Delta(u))| \ge \kappa\Delta\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) \\
&\le \mathbb{P}\Big(|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \ge |\varphi_\Delta(u)|\kappa\Delta\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) + \mathbb{P}\big(\Omega^c_{\zeta,\Delta}((n\Delta)^\alpha)\big).
\end{aligned}$$

Define $c(\Delta) := \kappa\Delta e^{-2\Delta}$, then, we derive from the Hoeffding inequality and Lemma 5.1 that

$$\mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \ge \kappa\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) \le 2(n\Delta)^{-c(\Delta)^2} + \frac{C}{(n\Delta)^2} + 4(n\Delta)^{\alpha-\eta},$$

where C depends on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$. Fix $\eta > 3$, such that $2\alpha - \eta < -1$ and $\zeta > \sqrt{\Delta(1+2\eta)}$, it follows that

$$T_n \le 4(n\Delta)^{\alpha - c(\Delta)^2} + \frac{C'}{n\Delta},$$

where C' depends on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$. Putting the above together, we have shown that for a positive constant C_1, depending on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$, and C_2 a constant depending on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$

$$\begin{aligned}
\mathbb{E}[\|\overline{f}_{\widehat{m}_n,\Delta} - f\|^2] \le C_1\Big(\|f_m - f\|^2 + \frac{\log(n\Delta)m}{n\Delta} + \frac{1}{n\Delta}\int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + \frac{m^2}{(n\Delta)^2}\Big) \\
+ C_2\Big(\frac{1}{n\Delta} + (n\Delta)^{\alpha - c(\Delta)^2}\Big).
\end{aligned}$$

Taking the infimum in m completes the proof. □

References

[1] A. Barron, L. Birgé, and P. Massart. "Risk bounds for model selection via penalization". In: *Probability theory and related fields* 113.3 (1999), pp. 301–413.

[2] J.-P. Baudry, C. Maugis, and B. Michel. "Slope heuristics: overview and implementation". In: *Statistics and Computing* 22.2 (2012), pp. 455–470.

[3] M. Bec and C. Lacour. "Adaptive pointwise estimation for pure jump Lévy processes". In: *Statistical Inference for Stochastic Processes* 18.3 (2015), pp. 229–256.

[4] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, and M. Reiß. "Lévy Matters IV". In: (2015).

[5] L. Birgé, P. Massart, et al. "Minimum contrast estimators on sieves: exponential bounds and rates of convergence". In: *Bernoulli* 4.3 (1998), pp. 329–375.

[6] B. Buchmann and R. Grübel. "Decompounding: an estimation problem for Poisson random sums". In: *Ann. Statist.* 31.4 (2003), pp. 1054–1074.

[7] C. Butucea and A. B. Tsybakov. "Sharp optimality in density deconvolution with dominating bias. I". In: *Teor. Veroyatn. Primen.* 52.1 (2007), pp. 111–128.

[8] C. Butucea and A. B. Tsybakov. "Sharp optimality in density deconvolution with dominating bias. II". In: *Teor. Veroyatn. Primen.* 52.2 (2007), pp. 336–349.

[9] C. Butucea. "Deconvolution of supersmooth densities with smooth noise". In: *Canadian Journal of Statistics* 32.2 (2004), pp. 181–192.

[10] R. J. Carroll and P. Hall. "Optimal rates of convergence for deconvolving a density". In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1184–1186.

[11] A. J. Coca. "Efficient nonparametric inference for discretely observed compound Poisson processes". In: *Probability Theory and Related Fields* (2017), pp. 1–49.

[12] F. Comte, Y. Rozenholc, and M.-L. Taupin. "Finite sample penalization in adaptive density deconvolution". In: *J. Stat. Comput. Simul.* 77.11-12 (2007), pp. 977–1000.

[13] F. Comte, C. Duval, and V. Genon-Catalot. "Nonparametric density estimation in compound Poisson processes using convolution power estimators". In: *Metrika* 77.1 (2014), pp. 163–183.

[14] F. Comte and V. Genon-Catalot. "Nonparametric adaptive estimation for pure jump Lévy processes". In: *Annales de l'institut Henri Poincaré (B)*. Vol. 46. 3. 2010, pp. 595–617.

[15] F. Comte and J. Kappus. "Density deconvolution from repeated measurements without symmetry assumption on the errors". In: *Journal of Multivariate Analysis* 140 (2015), pp. 31–46.

[16] F. Comte and C. Lacour. "Anisotropic adaptive kernel deconvolution". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 49. 2. Institut Henri Poincaré. 2013, pp. 569–609.

[17] F. Comte and C. Lacour. "Data-driven density estimation in the presence of additive noise with unknown distribution". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.4 (2011), pp. 601–627.

[18] F. Comte, Y. Rozenholc, and M.-L. Taupin. "Finite sample penalization in adaptive density deconvolution". In: *Journal of Statistical Computation and Simulation* 77.11 (2007), pp. 977–1000.

[19] A. Delaigle and I. Gijbels. "Practical bandwidth selection in deconvolution kernel density estimation". In: *Computational statistics & data analysis* 45.2 (2004), pp. 249–267.

[20] A. Delaigle, P. Hall, and A. Meister. “On deconvolution with repeated measurements”. In: *Ann. Statist.* 36.2 (2008), pp. 665–685.

[21] S. Delattre, M. Hoffmann, D. Picard, and T. Vareschi. “Blockwise SVD with error in the operator and application to blind deconvolution”. In: *Electronic journal of statistics* 6 (2012), pp. 2274–2308.

[22] D. L. Donoho and I. M. Johnstone. “Ideal spatial adaptation by wavelet shrinkage”. In: *biometrika* (1994), pp. 425–455.

[23] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. “Wavelet shrinkage: asymptopia?” In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1995), pp. 301–369.

[24] C. Duval. “Density estimation for compound Poisson processes from discrete data”. In: *Stochastic Process. Appl.* 123.11 (2013), pp. 3963–3986.

[25] C. Duval. “When is it no longer possible to estimate a compound Poisson process?” In: *Electronic journal of statistics* 8.1 (2014), pp. 274–301.

[26] C. Duval and J. Kappus. “Nonparametric adaptive estimation for grouped data”. In: *Journal of Statistical Planning and Inference* 182 (2017), pp. 12–28.

[27] B. van Es, S. Gugushvili, and P. Spreij. “A kernel type nonparametric density estimator for decompounding”. In: *Bernoulli* 13.3 (2007), pp. 672–694.

[28] J. Fan. “On the optimal rates of convergence for nonparametric deconvolution problems”. In: *The Annals of Statistics* (1991), pp. 1257–1272.

[29] A. Goldenshluger and O. Lepski. “Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality”. In: *The Annals of Statistics* (2011), pp. 1608–1632.

[30] A. Goldenshluger and O. Lepski. “On adaptive minimax density estimation on R^ d”. In: *Probability Theory and Related Fields* 159.3-4 (2014), pp. 479–543.

[31] A. Goldenshluger and O. Lepski. “Universal pointwise selection rule in multivariate function estimation”. In: *Bernoulli* 14.4 (2008), pp. 1150–1190.

[32] A. Goldenshluger and O. Lepski. “General selection rule from a family of linear estimators”. In: *Theory of Probability & Its Applications* 57.2 (2013), pp. 209–226.

[33] S. Gugushvili. “Nonparametric inference for discretely sampled Lévy processes”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 48. 1. Institut Henri Poincaré. 2012, pp. 282–307.

[34] S. Gugushvili, F. van der Meulen, and P. Spreij. “Nonparametric Bayesian inference for multidimensional compound Poisson processes”. In: *arXiv preprint arXiv:1412.7739* (2014).

[35] J. Johannes. “Deconvolution with unknown error distribution”. In: *The Annals of Statistics* 37.5A (2009), pp. 2301–2323.

[36] J. Johannes and M. Schwarz. “Adaptive circular deconvolution by model selection under unknown error distribution”. In: *Bernoulli* 19.5A (2013), pp. 1576–1611.

[37] J. Kappus. “Adaptive nonparametric estimation for Lévy processes observed at low frequency”. In: *Stochastic Process. Appl.* 124.1 (2014), pp. 730–758.

[38] J. Kappus and G. Mabon. “Adaptive density estimation in deconvolution problems with unknown error distribution”. In: *Electronic journal of statistics* 8.2 (2014), pp. 2879–2904.

[39] C. Lacour, P. Massart, and V. Rivoirard. “Estimator selection: a new method with applications to kernel density estimation”. In: *Sankhya A* 79.2 (2017), pp. 298–335.

[40] O. Lepskii. “Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates”. In: *Theory of Probability & Its Applications* 36.4 (1992), pp. 682–697.

[41] O. Lepskii. “On a problem of adaptive estimation in Gaussian white noise”. In: *Theory of Probability & Its Applications* 35.3 (1991), pp. 454–466.

[42] M. Lerasle. “Optimal model selection in density estimation”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 48. 3. Institut Henri Poincaré. 2012, pp. 884–908.

[43] K Lounici and R Nickl. “Uniform Risk Bounds and Confidence Bands in Wavelet Deconvolution”. In: *Annals of Statistics* 39 (2011), pp. 201–231.

[44] P. Massart. *Concentration inequalities and model selection*. Vol. 6. Springer, 2007.

[45] A. Meister. *Deconvolution problems in nonparametric statistics*. Berlin: Springer, 2009.

[46] A. Meister. “Optimal convergence rates for density estimation from grouped data”. In: *Statistics & probability letters* 77.11 (2007), pp. 1091–1097.

[47] M. H. Neumann. “On the effect of estimating the error density in nonparametric deconvolution”. In: *J. Nonparametr. Statist.* 7.4 (1997), pp. 307–330.

[48] M. H. Neumann and M. Reiß. “Nonparametric estimation for Lévy processes from low-frequency observations”. In: *Bernoulli* 15.1 (2009), pp. 223–248.

[49] M. Pensky, B. Vidakovic, et al. “Adaptive wavelet estimator for nonparametric density deconvolution”. In: *The Annals of Statistics* 27.6 (1999), pp. 2033–2053.

[50] G. Rebelles. “Structural adaptive deconvolution under {\mathbb {L} _p} -losses”. In: *Mathematical Methods of Statistics* 25.1 (2016), pp. 26–53.

[51] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. “Adaptive density estimation: a curse of support?” In: *Journal of Statistical Planning and Inference* 141.1 (2011), pp. 115–139.

[52] L. A. Stefanski. “Rates of convergence of some estimators in a class of deconvolution problems”. In: *Statistics & Probability Letters* 9.3 (1990), pp. 229–235.

[53] L. A. Stefanski and R. J. Carroll. “Deconvolving kernel density estimators”. In: *Statistics* 21.2 (1990), pp. 169–184.