Nonparametric adaptive estimation for grouped data

Céline Duval, Johanna Kappus

Nonparametric adaptive estimation for grouped data

C. Duval,[*] J. Kappus[†]

Abstract

The aim of this paper is to estimate the density f of a random variable X when one has access to independent observations of the sum of $K \geq 2$ independent copies of X. We provide a constructive estimator based on a suitable definition of the logarithm of the empirical characteristic function. We propose a new strategy for the data driven choice of the cut-off parameter. The adaptive estimator is proven to be minimax-optimal up to some logarithmic loss. A numerical study illustrates the performances of the method. Moreover, we discuss the fact that the definition of the estimator applies in a wider context than the one considered here.

Keywords. Convolution. Inverse problem. Nonparametric adaptive estimation.
AMS Classification. 62G07, 62G20, 62G05.

1 Introduction

1.1 Model and motivation

In this article, we consider the problem of nonparametric density estimation from grouped data observations. The model can be described as follows. Let X be some real-valued random variable which has a square integrable Lebesgue density f. We aim at estimating f, but we only have access to the aggregated data

$$Y_j = \sum_{k=1}^{K} X_{j,k}, \quad j = 1, \dots, n. \tag{1.1}$$

The $X_{j,k}$, $k = 1, \dots, K$; $j = 1, \dots, n$ are independent copies of X and K is a known positive integer.

This statistical framework is closely related to several other models. First, we have a connection to decompounding problems (see e.g. Buchmann and Grübel (2003) or van Es et al. (2007)), where the jump distribution is estimated from a discretely observed trajectory. In that latter example the law of each nonzero increment is an aggregation of several jumps.

Second, for $K \geq 2$, (1.1) describes a particular deconvolution problem where the target density is f and the error density is f^{*K-1}, which is unknown and depends on f itself. The classical deconvolution model, where the error distribution and the target distribution are not related,

---

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

has been extensively studied in the literature. Optimal rates of convergence and adaptive procedures are well known. In the case where the density of the noise is known we can mention Carroll and Hall (1988), Stefanski (1990), Stefanski and Carroll (1990), Fan (1991), Butucea (2004), Butucea and Tsybakov (2008a,b), Pensky and Vidakovic (1999) or Comte et al. (2007) for L_2 loss functions or Lounici and Nickl (2011) for the L_∞ loss. Deconvolution with unknown error distribution has also been studied (see e.g. Neumann (1997), Delaigle et al. (2008), Johannes (2009) or Meister (2009), if an additional error sample is available, or Comte and Lacour (2011), Delattre et al. (2012), Johannes and Schwarz (2013), Kappus and Mabon (2014) or Comte and Kappus (2015) under other sets of assumptions).

In the aforementioned literature, it is always assumed that the distribution of the noise is known or can be estimated from an additional sample. In the present framework, one has to exploit the particular structure of the observations and the dependence between the noise and the target density, in order to identify the latter one. Surprisingly, we will show that the optimal rates of convergence in the grouped data setting with skewed density are, in some cases, slower than the optimal rates which are usually encountered in deconvolution problems (see Theorem 2 below).

Recently, Delaigle and Zhou (2015) studied the nonparametric estimation of the probability of contamination given a covariate X, when only aggregated observations of this covariate are available. From an applied point of view, grouped data are encountered in many fields, especially related to the study of infection diseases. For example, aggregated data are encountered when the cost of measurement is too high to allow individual measurements. Moreover, to ensure confidentiality of the participants in a study, measurements of different individuals can be grouped. Sometimes, when X represents the measurement of a substance that cannot be detected below a certain detection limit, observations are grouped in order to make the detection possible (see the references given in Delaigle and Zhou (2015)).

1.2 Main results of the paper

We have at our disposal n independent copies of the random variable Y, which has the Lebesgue density f^{*K}, where $*$ denotes the convolution product. This implies that $\varphi_Y(u) = \varphi_X(u)^K$, where φ_X and φ_Y denote the characteristic functions of X and Y respectively. Consequently, f can be recovered, taking the K-th root of the characteristic function of φ_Y and applying the Fourier inversion formula. A first estimator based on this strategy is proposed in Linton and Whang (2002) to estimate a regression function when only aggregated observations are available. They studied the asymptotic pointwise behavior of their estimator.

If φ_X has no imaginary part and is nonnegative, the K-th root is uniquely defined and an estimator of φ_X can hence be constructed, replacing φ_Y by the absolute value of its empirical counterpart and taking the K-th root. An estimator of f is then obtained applying a Fourier inversion formula. Meister (2007) studied this case and provided the associated minimax upper and lower bounds. Moreover, a cross-validation bandwidth selector is proposed in that paper. However, theoretical optimality properties for the adaptive procedure have only been shown under quite restrictive assumptions on the characteristic function.

Meister (2007) also proposed a generalization of the estimation procedure to the skewed case

and a modified version of this estimator is used by Delaigle and Zhou (2015). However, this estimator has the disadvantage that it is not given in a closed form, making the practical calculation difficult to handle.

In the skewed case, the estimation procedure suffers from the fact that the definition of K-th root is ambiguous. In the present paper we provide a definition of the K-th root, based on a suitable definition of the logarithm (see Lemma 3 hereafter), of a characteristic function that is not necessarily real. A closed form procedure to estimate f naturally follows. For this estimator, we provide non asymptotic risk bounds and derive rates of convergence over certain nonparametric function classes.

It is interesting to note that the smoothness classes considered in Meister (2007), which differ from usual Sobolev and Hölder classes, are not very general, since the characteristic functions are bounded from above and below by functions which are of the same order. In the present work, we provide rates of convergence over more general function classes. In the particular cases which have been considered in Meister (2007), our rates of convergence agree with the ones found in that paper.

However, surprisingly, in the general case our rates of convergence are different from the rates which are known to be optimal in a classical deconvolution framework, where the noise density is known or can be estimated from an additional sample. We prove that these unusual rates are optimal in the grouped data framework (see Theorem 2). This reveals that the grouped data setting has its own specificities and differ from the classical deconvolution framework. To the best of our knowledge, this phenomenon is completely new in the literature on deconvolution.

Finally, we propose a new data driven selection strategy for the cut-off parameter. We show that the resulting rates of convergence are, up to a logarithmic loss, adaptive minimax. Our adaptive procedure is computationally simple and extremely fast, in comparison to usual cross-validation techniques. Moreover, it applies in a wider setting than the one proposed in Meister (2007).

1.3   Organization and some notation

The paper is organized as follows. Section 2 contains the construction of the estimator. Upper and lower risk bounds are established and the corresponding minimax rates are derived. In Section 3 a data driven bandwidth selector is introduced and it is shown that the resulting rates are adaptive minimax. A discussion is proposed in Section 4 and a numerical study is given in Section 5. Finally, the proofs are gathered in Section 6.

We close this section by introducing some notation which will be used throughout the rest of the text. In the sequel, $*$ is understood to be the convolution product of integrable functions and f^{*K} denotes the K-fold auto-convolution, $f^{*K} = f * \ldots * f$. Given a random variable Z, $\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$ denotes the characteristic function of Z. For Y_j defined as in formula (1.1), we will drop the subscript and write φ instead of φ_{Y_j}. 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. 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 $\mathcal{F}f_m = (\mathcal{F}f)\mathbb{1}_{[-m,m]}$.

2 Risk bounds and rates of convergence

2.1 Construction of the estimator

In the sequel, we will always work under the assumption

$$\forall u \in \mathbb{R}, \quad \varphi_X(u) \neq 0. \tag{A0}$$

This assumption is standard in deconvolution problems and, for the present model, proven necessary in Meister (2007). Therein, a counterexample is made explicit. When K is even and (A0) is not satisfied, it is possible to build two densities f_0 and f_1 which differ on a set of nonzero Lebesgue measure and for which $f_0^{*K}(x) = f_1^{*K}(x)$, $\forall x \in \mathbb{R}$. Hence it is not possible to distinguish f_0 from f_1 in presence of the K aggregated observations.

Suppose, we are given n independent copies of the random variable $Y = \sum_{k=1}^{K} X_{1,k}$, which has the Lebesgue density f^{*K}. In the Fourier domain, this implies $\varphi(u) = \varphi_X(u)^K$. Consequently, f can be recovered, taking the K-th root of the characteristic function and using the fact that, by the Fourier inversion formula,

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \big(\varphi(u)\big)^{1/K} \, \mathrm{d}u.$$

This formula makes immediate sense when φ has no imaginary part and is strictly positive. Contrarily, if f is non symmetric and hence φ has a nonzero imaginary part, the definition of $z \mapsto z^{1/K}$ is not unique. In order to get a proper definition of the K-th root, one needs to specify which branch of the logarithm is considered. The algorithm discussed in Meister (2007) for the non symmetric case is not given in a closed form so the estimator is numerically difficult to handle.

In order to get a closed-form estimator let us recall the concept of the distinguished logarithm.

Lemma 1. *Let φ be a characteristic function which has no zeros. Then there exists a unique continuous function ψ which satisfies*

1. $\psi(0) = 0$

2. $\forall u \in \mathbb{R} : \varphi(u) = e^{\psi(u)}$.

The function ψ introduced in Lemma 1 is called the *distinguished logarithm* of φ and, hereafter, denoted by $\mathrm{Log}\,\varphi$. Moreover, in the sequel, we let $1/K$ denote the *distinguished K-th* root, that is, $\varphi^{1/K} := \exp((\mathrm{Log}\,\varphi)/K)$. It is important to keep in mind that the distinguished logarithm is usually not equal to the main branch of the logarithm. Moreover, the definition of a distinguished logarithm, and consequently of a distinguished root, only makes sense with respect to φ seen as a function rather than pointwise.

Lemma 2. *For some integer K, let $\varphi(u) = \varphi_X(u)^K$. Then, both characteristic functions are related as follows,*

$$\forall u \in \mathbb{R} : \varphi_X(u) = \exp((\mathrm{Log}\,\varphi(u))/K) =: \varphi(u)^{\frac{1}{K}}.$$

The proof of Lemma 1 and Lemma 2 can be found, for example, in Sato (1999) (Lemma 7.6). In order to obtain a constructive estimator for φ_X, we give the following result which makes the definition of the distinguished logarithm explicit.

Lemma 3. *Let φ be a characteristic function without zeroes and assume that φ is differentiable. Then, it follows that*

$$\text{Log}\,(\varphi(u)) = \int_0^u \frac{\varphi'(z)}{\varphi(z)}\,\mathrm{d}z. \tag{2.2}$$

The preceding lemmas suggest to exploit formula (2.2) in order to build a constructive estimator of the distinguished logarithm and hence of the characteristic function itself. The characteristic function of Y and its derivative are replaced by their empirical counterparts,

$$\widehat{\varphi}(u) := \frac{1}{n}\sum_{j=1}^{n} e^{iuY_j} \quad \text{and} \quad \widehat{\varphi}'(u) = \frac{1}{n}\sum_{j=1}^{n} iY_j e^{iuY_j}.$$

Moreover, we define the quantities

$$\psi'(u) := \frac{\varphi'(u)}{\varphi(u)} \quad \text{and} \quad \widehat{\psi}'(x) := \frac{\widehat{\varphi}'(u)}{\widehat{\varphi}(u)}.$$

as well as

$$\psi(u) := \int_0^u \psi'(x)\,\mathrm{d}x \quad \text{and} \quad \widehat{\psi}(u) := \int_0^u \widehat{\psi}'(x)\,\mathrm{d}x.$$

It follows from Lemma 3 that $\psi(u) = \text{Log}\,(\varphi(u))$ and $\widehat{\psi}(u) = \text{Log}\,(\widehat{\varphi}(u))$. Lemma 2 suggests to use

$$\widehat{\varphi}_X(u) := \exp(\widehat{\psi}(u)/K)$$

as an estimator of φ_X. Finally, we use a spectral cut-off and apply Fourier inversion to derive, for arbitrary $m > 0$, the following estimator of f,

$$\widehat{f}_m(x) := \frac{1}{2\pi}\int_{-m}^{m} e^{-iux}\widehat{\varphi}_X(u)\,\mathrm{d}u. \tag{2.3}$$

It is worth pointing out that for real-valued and strictly positive characteristic functions, our estimator coincides with the estimator given in Meister (2007). In this case, the distinguished K-th root equals the usual K-th root. This estimator resembles one of the estimators defined in Comte et al. (2015), who consider the problem of estimating the jump density for mixed compound Poisson processes. However, in Comte et al. (2015) the associated upper bound is sub optimal and an adaptive bandwidth selector is not proposed.

2.2 Non asymptotic risk bounds

For $\varepsilon, \gamma > 0$, define

$$u_n^{(\gamma,\varepsilon)} := \min\{u \geq 0 : |\varphi(u)| = (1+\varepsilon)\gamma(n/\log n)^{-1/2}\}. \tag{2.4}$$

The following bound is valid for the mean integrated squared error of $\widehat{f}_m$.

Theorem 1. *Assume that X has a finite second moment. Let $\gamma := \sqrt{1 + \frac{2}{K} + \delta}$, for some $\delta > 0$. Then, for any $m \geq 1$,*

$$\mathbb{E}[\|f - \widehat{f}_m\|^2] \leq \|f - f_m\|^2 + \frac{C_1}{nK^2} \int\limits_{-m}^{m} \frac{1}{|\varphi_X(u)|^{2(K-1)}} \, \mathrm{d}u + C_2 mn^{-\frac{1}{K}} + \frac{2}{\pi} \int\limits_{\{|u| \in U_n^{\gamma,\varepsilon}\}} |\varphi_X(u)|^2 \, \mathrm{d}u, \quad (2.5)$$

where $U_n^{\gamma,\varepsilon} = [u_n^{(\gamma,\epsilon)}, m]$, with the convention that if $m \leq u_n^{(\gamma,\epsilon)}$ then $U_n^{\gamma,\varepsilon} = \emptyset$, and where $C_1 = (1 + \varepsilon) \log(1 + 1/\varepsilon)$ and $C_2 = (9 + 4\|\varphi_X\|^2)$.

The first two terms appearing in the upper risk bound illustrate the structural similarity between this grouped data setting and a classical density deconvolution framework. The density f of X plays the role of the density of interest and $f^{*(K-1)}$ resembles the density of the error term. The third summand corresponds to the usual variance term for density estimation from grouped data when the density is symmetric (see Meister (2007)). Finally, the fourth summand depends on the quantity $u_n^{(\gamma,\epsilon)}$ which appears naturally. We prove that on the interval $[-u_n^{(\gamma,\epsilon)}, u_n^{(\gamma,\epsilon)}]$, the estimator $\widehat{\varphi}$ can be controlled uniformly. This allows to control the deviation of the empirical distinguished logarithm and consequently the deviation of $\widehat{\varphi}_X$ from φ_X.

It follows directly from the proof of Theorem 1 (see Section 6.2) that in the symmetric case the upper bound simplifies, for any $m > 0$, to

$$\mathbb{E}[\|f - \widehat{f}_m\|^2] \leq \|f - f_m\|^2 + Cmn^{-\frac{1}{K}}.$$

The second and fourth term appearing on the right hand side of formula (2.5) are particular to the case where the density is skewed. Balancing those four terms, there will occur a logarithmic loss, compared to the symmetric case. This phenomenon is a consequence of the fact that the estimation problem in the non-symmetric case involves the distinguished logarithm, which requires uniform rather than pointwise control of the empirical characteristic function. Controlling suprema typically leads to the loss of a logarithmic factor. However, we do not have a lower bound result that establishes whether this logarithmic loss is avoidable or not.

2.3 Rates of convergence

Let us investigate the rates of convergence over certain nonparametric classes of densities. In the sequel, for $b \geq \beta > \frac{1}{2}$, we denote by $\mathcal{F}(\beta, b, C, C', C_X, C_f)$ the class of densities for which the following holds

$$\forall u \in \mathbb{R} : (1 + C|u|)^{-b} \leq |\varphi_X(u)| \leq (1 + C'|u|)^{-\beta} \tag{2.6}$$

and, in addition,

$$\mathbb{E}[X^2] \leq C_X \quad \text{and} \quad \|f\| \leq C_f. \tag{2.7}$$

6

For $\rho > 0$, $\beta > \frac{1}{2}$, $b \in \mathbb{R}$ and $c > 0$, let $\mathcal{H}(\beta, \rho, b, c, C, C', C_X, C_f)$ be the class of densities for which (2.7) holds and, in addition,

$$\forall u \in \mathbb{R} : (1 + C|u|)^{-b} \exp(-c|u|^\rho) \le |\varphi_X(u)| \le (1 + C'|u|)^{-\beta}.$$

Finally, we denote by $\mathcal{G}(\rho, c, C, C', C_X, C_f)$ the class of densities for which (2.7) holds and, in addition,

$$\forall u \in \mathbb{R} : (1 + C|u|)^{(\rho-1)/2} \exp(-c|u|^\rho) \le |\varphi_X(u)| \le (1 + C'|u|)^{(\rho-1)/2} \exp(-c|u|^\rho). \quad (2.8)$$

In what follows, for the sake of readability, we drop the dependence on C, C', C_X and C_f when there is no risk of confusion. We write, for example $\mathcal{F}(\beta, b)$ instead of $\mathcal{F}(\beta, b, C, C', C_X, C_f)$. As it is usually employed in the literature, we say that a density is "ordinary smooth" if its characteristic function decays polynomially as in (2.6) and that it is "super smooth" if its characteristic function decays exponentially as in (2.8).

The classes of functions introduced above differ from the Sobolev or Hölder classes which are usually considered in density estimation. These types of function classes are characterized by some bound from above on the decay of the characteristic function. But contrary to the present model, bounds from below are not imposed. However, the statistical model under consideration is a special case of a deconvolution problem, with target density f and noise density $f^{*(K-1)}$. In a deconvolution framework, rates of convergence over nonparametric classes of target densities can only be established under additional assumptions on the decay of the characteristic function of the noise. This explains why we consider classes of functions that provide a control from above and below of the characteristic function. It permits to render explicit the rates of convergence, and particular to be able to handle the second term in (2.5) and the quantity $u_n^{\gamma, \varepsilon}$ defined in (2.4).

The following result is an immediate consequence of Theorem 1.

Proposition 1. *For $a > 0$, set*

$$u_n^{(a)} := \min\{u \ge 0 : |\varphi(u)| = an^{-\frac{1}{2}}\}.$$

Fix $a, \varepsilon, \delta > 0$ and let γ be defined as in Theorem 1. For every $n \in \mathbb{N}$, let $m^* \in [u_n^{(\gamma, \epsilon)}, u_n^{(a)}]$. Then, we have

(i) $\quad \displaystyle\sup_{f \in \mathcal{F}(\beta, b)} \mathbb{E}[\|f - \widehat{f}_{m^*}\|^2] = O\left((n/\log n)^{-\frac{(2\beta-1)}{2Kb}}\right).$

(ii) $\quad \displaystyle\sup_{f \in \mathcal{H}(\beta, \rho, b, c)} \mathbb{E}[\|f - \widehat{f}_{m^*}\|^2] = O\left((\log n)^{\frac{2\beta-1}{\rho}}\right).$

(iii) $\quad \displaystyle\sup_{f \in \mathcal{G}(\rho, c)} \mathbb{E}_f[\|f - \widehat{f}_{m^*}\|^2] = O\left(n^{-\frac{1}{K}} (\log n)^{\frac{1}{\rho}}\right).$

If we consider the class $\mathcal{F}(\beta,\beta)$ (taking $b = \beta$) or the class $\mathcal{G}(\rho,c)$, the rates of convergence summarized in Proposition 1 coincide with the rates which have been proven to be minimax optimal (up to a logarithmic loss) in Meister (2007). Moreover, the rates of convergence derived for $\mathcal{H}(\beta,\rho,b,c)$ coincide with the optimal rates of convergence encountered in a deconvolution setting when the target density is ordinary smooth and the density of the noise is super smooth, see e.g. Fan (1991).

The general class $\mathcal{F}(\beta,b)$, with $b > \beta$, has not been investigated in Meister (2007). It is unexpected and noteworthy that the rate of convergence derived in Proposition 1 differs from the optimal rate of convergence which is "standard" in deconvolution problems.
Indeed, take the case where a random variable X is observed with an additional additive error ϵ whose distribution is known. If $|\varphi_X|$ is bounded from above by a function which decays as $|u|^{-\beta}$ and $|\varphi_\epsilon|$ is bounded from below by a function decaying as $|u|^{-(K-1)b}$, the optimal rate of convergence is $n^{-\frac{2\beta-1}{2\beta+2(K-1)b}}$ (see e.g. Meister (2009)). If the characteristic function of ϵ is unknown, but can be estimated from an additional sample of size n of the pure noise, the rate of convergence is the same as in the case of a known error distribution (see e.g. Johannes (2009)).
In comparison to this, the rate of convergence derived in Proposition 1 equals, up to a logarithmic factor $n^{\frac{2\beta-1}{2bK}}$, which is slower. It is shown in the next section that this slower rate of convergence is not due to our estimator, but is, up to the logarithmic factor, optimal in the grouped data framework.

We have not discussed the case where $|\varphi_X|$ is bounded from above and below by exponentially decaying functions with different orders. It is possible to derive rates of convergence in this case. However, computations are cumbersome and the resulting rates are, in some particular cases, not available in a closed form (see Lacour (2006)). For this reason, we omit this case.

2.4 Minimax lower bound

The following result shows that, if $\beta > 1$, the rates of convergence found for the class $\mathcal{F}(\beta,b)$ are minimax optimal up to the loss of a logarithmic factor.

Theorem 2. *Assume that $\beta > 1$. Then there exists some positive constant d such that*

$$\liminf_{n\to\infty} \inf_{\widetilde{f}} \sup_{f\in\mathcal{F}(\beta,b)} \mathbb{E}_f[\|f - \widetilde{f}\|^2] n^{\frac{2\beta-1}{2bK}} \geq d > 0.$$

The infimum is taken over all estimators $\widetilde{f}$ of f, based on the observations Y_j, $j = 1,\ldots,n$.

Theorem 2 does not cover the cases where $\beta \in (0.5, 1]$. However, at the expense of some additional technicalities, it is possible to extend the result to this case. We neither formulate nor prove lower bound results associated to the upper bounds of Proposition 1, part (ii) and (iii). Indeed, part (ii) matches the optimal convergence rate for density deconvolution problems with known error density and part (iii) has been proven optimal in Meister (2007).

8

Gathering the results of Proposition 1 and Theorem 2 we have established that our estimation procedure is rate optimal up to a logarithmic factor. It remains unclear if the logarithmic loss is avoidable. However, we suspect that the rate cannot be improved as some uniform control of the empirical characteristic function is required, which typically leads to a logarithmic loss.

3   Adaptive estimation

The cut-off m^* is defined in terms of the characteristic function: the estimator of $\widehat{\varphi}_X$ is set to zero as soon as $\varphi(u)$ is below some critical threshold which is essentially of the order $n^{-1/2}$. This is intuitive, since a reasonable estimate of the characteristic function in the denominator is no longer possible when this object is of smaller order than the standard deviation. It is interesting to notice that this theoretical cut-off m^*, which may vary in a certain interval and guarantees rate optimality of the estimator, has an empirically accessible counterpart. This motivates the definition of the empirical cut-off in terms of the empirical characteristic function. Once $\widehat{\varphi}$ is (up to an additional logarithmic term) below some critical threshold, $\widehat{\varphi}_X$ is set to zero. For a numerical constant $\eta > 1$, to be chosen, we introduce the data driven version of m^*,

$$\widehat{m}_\eta := \min \Big\{ \min\{u : |\widehat{\varphi}(u)| \leq (Kn)^{-\frac{1}{2}} + \sqrt{\eta/K}(n/\log n)^{-\frac{1}{2}}\}, n^{\frac{1}{K}}\Big\}.$$

The corresponding estimator $\widehat{f}_{\widehat{m}_\eta}$ adapts automatically to the unknown smoothness class, so the convergence rate is simultaneously minimax over a collection of nonparametric classes.

Theorem 3. *Let constants* $\overline{B}, \underline{B} > \frac{1}{2}, \overline{C}, \underline{C}' > 0, \overline{C}_X$ *and* $\overline{C}_f$ *be given.*

(i) Define

$$I := \{(\beta, b) : \overline{B} \geq b \geq \beta > \underline{B}\} \times (0, \overline{C}] \times [\underline{C}', \infty) \times (0, \overline{C}_X] \times (0, \overline{C}_f].$$

Then, there exists some positive real $\mathcal{C}$ *such that*

$$\sup_{(\beta, b, C, C', C_X, C_f) \in I} \left(\frac{\sup_{f \in \mathcal{F}(\beta, b, C, C', C_X, C_f)} \mathbb{E}_f[\|f - \widehat{f}_{\widehat{m}_\eta}\|^2]}{(n/\log n)^{-\frac{(2\beta - 1)}{2Kb}}} \right) \leq \mathcal{C}.$$

(ii) Define

$$I := [\underline{B}, \infty) \times (0, \overline{\rho}] \times (-\infty, \overline{B}] \times (0, \overline{c}] \times (0, \overline{C}] \times [\underline{C}', \infty) \times (0, \overline{C}_X] \times (0, \overline{C}_f].$$

Then, there exists a positive real $\mathcal{C}$ *such that*

$$\sup_{(\beta, \rho, b, c, C, C', C_X, C_f) \in I} \left(\frac{\sup_{f \in \mathcal{H}(\beta, \rho, b, c, C, C', C_X, C_f)} \mathbb{E}_f[\|f - \widehat{f}_{\widehat{m}_\eta}\|^2]}{(\log n)^{-\frac{(2\beta - 1)}{\rho}}} \right) \leq \mathcal{C}.$$

(iii) Let

$$I = (0, \overline{\rho}] \times (0, \overline{c}] \times [\underline{C}, \overline{C}] \times [\underline{C}', \overline{C}'] \times (0, \overline{C}_X] \times (0, \overline{C}_f].$$

Then, there exists some positive real $\mathcal{C}$ *such that*

$$\sup_{(\rho, c, C, C', C_X, C_f) \in I} \left(\frac{\sup_{f \in \mathcal{G}(\rho, c, C, C', C_X, C_f)} \mathbb{E}_f[\|f - \widehat{f}_{\widehat{m}_\eta}\|^2]}{n^{-\frac{1}{K}}(\log n)^{\frac{1}{\rho}}} \right) \leq \mathcal{C}.$$

In earlier works, the adaptive bandwidth selection has only been investigated for the particular case of symmetric densities and for polynomially decaying characteristic functions. Moreover, the cross-validation strategy proposed in Meister (2007) cannot be extended to the present framework, since it leads to the choice $n^{-1/(2\beta K)}$, which gives a sub-optimal rate of convergence for $b > \beta$. Finally, in comparison to cross validation, our adaptive estimator has the strong advantage of being computationally extremely simple and fast, since it only requires the evaluation of the characteristic function.

The adaptive choice $\widehat{m}_\eta$ depends on the constant η that needs to be chosen. I is not unusual when it comes to adaptation that an hyper parameter needs to be calibrated. The same holds true in the context of wavelet adaptive estimation or when the spectral cut-off is obtained by minimization of a penalty. Our simulation results indicate that the procedure works well if η is chosen close to 1.

4   Discussion

In this article, we provided a fully constructive adaptive and optimal estimation procedure in the context of grouped data observations, which covers both the non symmetric and symmetric cases. The study of this framework revealed that it has its own specificities and differs from what one would expect in a classical deconvolution framework.
It is also worth noticing that if K increases, the rates of convergence obtained deteriorate in Theorem 3. If we let $K \to \infty$ then, (see Meister (2007) for the lower bound results)

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}} \mathbb{E}\big(\|\widehat{f}_m - f\|^2\big) > 0,$$

where $\mathcal{F}$ is a class of densities, for instance one of those defined in Section 2.3. This result is intuitive. Assume that the density f has expectation μ and finite variance σ^2. Then we have

$$\frac{(Y_j - \mu)}{\sqrt{K}} \underset{K \to \infty}{\longrightarrow} \mathcal{N}(0, \sigma^2).$$

It means that for large K, each observation Y_j is close in law to a Gaussian random variable depending on two parameters only. The whole density f is then lost, only its expectation and variance may be recovered from the observations. This phenomenon is also observed in Duval (2014) in the case of a compound Poisson process observed at a sampling rate tending to infinity.

Finally, we want to emphasize that the adaptive estimator we define in this paper applies for a wider range of problems than grouped data. In fact, it can be used each time the quantity of interest can be recovered through a closed form inverse of the characteristic function of the observations.
Suppose that a Lévy process Y is observed over $[0, T]$ at low frequency and denote by $\Delta \geq 1$ the sampling rate. A vast literature is available on the estimation of the underlying parameters, and in particular the estimation of the Lévy measure (see among others Neumann and Reiß (2009), Comte and Genon-Catalot (2010) or Kappus (2014)). However, one may also

consider the case where the quantity of interest is the distributional density of Y_t for some $t > 0$ (without loss of generality set $t = 1$) rather than the jump density.

Suppose that Y_1 has a square integrable Lebesgue density f. Let φ_1 and φ_Δ be the characteristic functions of Y_1 and Y_Δ, respectively. Then both characteristic functions are connected as $\varphi_\Delta(u) = (\varphi_1(u))^\Delta$ or equivalently as $\varphi_1 = \exp((\mathrm{Log}\,\varphi_\Delta)/\Delta)$. The estimation procedure proposed in this paper carries over immediately, leading to the estimator

$$\widehat{f}_m(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \widehat{\varphi}_1(u)\, \mathrm{d}u, \quad \text{with} \quad \widehat{\varphi}_1(u) = \exp\Big(\frac{1}{\Delta} \int_0^u \frac{\widehat{\varphi}'_\Delta(v)}{\widehat{\varphi}_\Delta(v)}\, \mathrm{d}v \Big)$$

and the upper bound

$$\mathbb{E}[\|f - \widehat{f}_m\|^2] \leq \|f - f_m\|^2 + \frac{C_1}{\Delta^2 n} \int_{-m}^{m} \frac{|\varphi_1(u)|^2}{|\varphi_\Delta(u)|^2}\, \mathrm{d}u + C_2 \int_{[m, u_n^{(\gamma, \epsilon)}]} |\varphi_1(u)|^2\, \mathrm{d}u + C_3 n^{-\frac{1}{\Delta}}.$$

Balancing the terms when φ_1 is polynomially decreasing leads to the rate $(T/\log T)^{-\frac{2\beta-1}{2\Delta\beta}}$. For exponentially decaying characteristic functions, the convergence rate is, up to a logarithmic term, $n^{-\frac{1}{\Delta}}$. In a sense, the context of infinitely divisible distributions can be interpreted as a "grouped data model" with a non-integer group size Δ. The choice of $\widehat{m}_\eta$ proposed in the present publication will lead to a minimax adaptive estimator.

5 Numerical results

In this section we illustrate the behavior of our adaptive procedure by computing L_2 risks for various density functions f, when n and K are varying. The upper bound of Theorem 1 states that these risks should decrease as n increases and should increase with K. As points of comparison, we compute also the risks of two different procedures. First, the cross-validation estimator proposed in Meister (2007), even though no theoretical justification has been established that the cross-validation strategy works outside the particular case where f is symmetric and the characteristic function decays polynomially. Second, an oracle "estimator". More precisely the estimator $\widehat{f}_{m^\star}$, which is the estimator defined in (2.3) where $m^\star$ corresponds to the following oracle bandwidth

$$m^\star = \arg\min_{m>0} \|f - \widehat{f}_m\|^2.$$

This oracle can be explicitly computed if f is known. We denote these different risks by r, for the risk of our procedure, r_{cv} for the cross-validation estimator and r_{or}^* for the oracle procedure. All three estimators are computed on the same samples.

The L^2 risks are computed after 500 Monte-Carlo simulations of each estimator and averaging the results. We consider the following group size $K = 5$, 10, 20 and 50 and the following distributions:

(i) Normal distribution with mean value 2 and variance 1.

(ii) Gumbel distribution with mean 3 and scaling parameter 1.

11

(iii) Gamma distribution with parameters 6 and 3.

(iv) Laplace distribution with mean 0.5 and scale parameter $\beta = 3$.

Moreover, we set $\eta = 1.1$ for the adaptive procedure. Numerical experiments indicate that the procedure is sensitive to the choice of the numerical constant η and that this constant should be chosen close to 1. The results are summarized in the tables below.

n	K	$\mathcal{N}(2,1)$			Gumbel(3,1)		
		r_{or}^*	r	r_{cv}	r_{or}^*	r	r_{cv}
1000	5	0.033	0.088	0.095	0.017	0.037	0.045
	10	0.045	0.124	0.144	0.031	0.066	0.082
	20	0.057	0.156	0.185	0.043	0.092	0.111
	50	0.069	0.185	0.236	0.053	0.117	0.154
5000	5	0.031	0.074	0.073	0.012	0.027	0.035
	10	0.039	0.114	0.123	0.027	0.056	0.062
	20	0.053	0.147	0.168	0.040	0.083	0.101
	50	0.066	0.183	0.227	0.052	0.110	0.143
10000	5	0.007	0.018	0.021	0.011	0.025	0.038
	10	0.017	0.045	0.055	0.025	0.053	0.059
	20	0.028	0.077	0.099	0.039	0.081	0.094
	50	0.039	0.105	0.221	0.051	0.114	0.140

n	K	$\Gamma(6,3)$			Laplace(0.5,3)		
		r_{or}^*	r	r_{cv}	r_{or}^*	r	r_{cv}
1000	5	0.021	0.050	0.060	0.070	0.152	0.149
	10	0.039	0.089	0.110	0.114	0.239	0.260
	20	0.054	0.125	0.158	0.150	0.312	0.367
	50	0.067	0.162	0.212	0.180	0.372	0.468
5000	5	0.016	0.037	0.050	0.055	0.118	0.121
	10	0.033	0.076	0.090	0.100	0.214	0.231
	20	0.049	0.116	0.134	0.140	0.294	0.324
	50	0.065	0.159	0.202	0.173	0.374	0.438
10000	5	0.013	0.032	0.040	0.047	0.103	0.098
	10	0.030	0.071	0.090	0.094	0.197	0.208
	20	0.047	0.111	0.131	0.134	0.287	0.319
	50	0.064	0.153	0.197	0.170	0.350	0.439

From the above tables, we observe that all procedures performs well in all examples. As expected, we observe that for all procedures, the associated risks decrease as n increases and that increasing K deteriorates the risks. Moreover the adaptive strategy proposed in the present article shows a slightly better outcome than the cross-validation techniques. It is also numerically extremely efficient since it requires only one evaluation of the empirical characteristic function.

6 Proofs

6.1 Preliminary results

Lemma 4. *Let Z be an integrable random variable with characteristic function φ. Then, for any $h > 0$,*
$$\forall u \in \mathbb{R} : |\varphi(u) - \varphi(u + h)| \leq h\mathbb{E}[|Z|].$$

Proof. Using that for any $x, y \in \mathbb{R}$, $|e^{ix} - e^{iy}| \leq |x - y|$, leads to the result
$$|\varphi(u) - \varphi(u + h)| = |\mathbb{E}[e^{iuZ} - e^{i(u+h)Z}]| \leq h\mathbb{E}[|Z|].$$
$\square$

For arbitrary $c > 0$, define the event
$$A_c := \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} (|Y_j| - \mathbb{E}[|Y|]) \right| \leq c \right\}.$$

By the Markov inequality, we derive that
$$\mathbb{P}(A_c^c) \leq \frac{1}{c^2 n} \mathbb{E}[Y^2].$$

Lemma 5. *Let $h > 0$. Then we have on the event A_c the following inequality*
$$\forall u \in \mathbb{R} : |\widehat{\varphi}(u) - \widehat{\varphi}(u + h)| \leq h(\mathbb{E}[|Y|] + c).$$

Proof. The triangle inequality and the fact that $x \to e^{iux}$ is Lipschitz with constant 1, give
$$|\widehat{\varphi}(u) - \widehat{\varphi}(u + h)| \leq \frac{h}{n} \sum_{j=1}^{n} |Y_j|.$$

The definition of A_c leads to
$$\frac{h}{n} \sum_{j=1}^{n} |Y_j| = h\left(\mathbb{E}[|Y|] + \frac{1}{n} \sum_{j=1}^{n} (|Y_j| - \mathbb{E}[|Y|]) \right) \leq h(\mathbb{E}[|Y|] + c).$$
$\square$

Lemma 6. *Fix $h > 0$ and some $u > 0$. Define the grid points $t_k := kh$, $k \in \{1, \ldots \lceil \frac{u}{h} \rceil\}$. Then the following holds true for arbitrary $\tau > 0$.*
$$\mathbb{P}\left(\exists k \in \{1, \ldots, \lceil \tfrac{u}{h} \rceil\} : |\widehat{\varphi}(t_k) - \varphi(t_k)| > \tau \left(\frac{\log n}{n} \right)^{\frac{1}{2}} \right) \leq 2\lceil \tfrac{u}{h} \rceil n^{-\tau^2/2}.$$

Proof. This is a consequence of the series of inequalities
$$\mathbb{P}\left(\exists k \leq \lceil \tfrac{u}{h} \rceil : |\widehat{\varphi}(t_k) - \varphi(t_k)| \geq \tau \left(\frac{\log n}{n} \right)^{\frac{1}{2}} \right) \leq \sum_{k=1}^{\lceil u/h \rceil} \mathbb{P}\left(|\widehat{\varphi}(t_k) - \varphi(t_k)| \geq \tau \left(\frac{\log n}{n} \right)^{\frac{1}{2}} \right)$$
$$\leq \sum_{k=1}^{\lceil u/h \rceil} \mathbb{P}\left(\sum_{j=1}^{n} |e^{it_k Y_j} - \mathbb{E}[e^{it_k Y_j}]| \geq \tau (n\log n)^{\frac{1}{2}} \right) \leq \sum_{k=1}^{\lceil u/h \rceil} 2\exp(-\tau^2(\log n)) = 2\lceil u/h \rceil n^{-\frac{\tau^2}{2}}.$$

The last inequality follows from the Hoeffding inequality. $\square$

Fix c, h and τ. For arbitrary $u > 0$, define the event

$$B_{c,h,\tau}(u) := A_c \cap \left\{ \forall k \leq \lceil \tfrac{u}{h} \rceil : |\widehat{\varphi}(t_k) - \varphi(t_k)| \leq \tau \left(\frac{\log n}{n} \right)^{1/2} \right\}.$$

Lemma 7. *On the event $B_{c,h,\tau}(u)$, the following holds*

$$\sup_{v \in [0,u]} |\widehat{\varphi}(v) - \varphi(v)| \leq h(2\mathbb{E}[|Y|] + c) + \tau \left(\frac{\log n}{n} \right)^{1/2}.$$

Proof. Consider the grid point t_k defined in Lemma 6 and $v \in [t_k - h/2, t_k + h/2]$ for some $k \leq \lceil \tfrac{u}{h} \rceil$. Then, applying Lemma 4 and Lemma 5, we have on the event $B_{c,h,\tau}(u)$

$$|\widehat{\varphi}(v) - \varphi(v)| \leq |\widehat{\varphi}(v) - \widehat{\varphi}(t_k)| + |\widehat{\varphi}(t_k) - \varphi(t_k)| + |\varphi(t_k) - \varphi(v)|$$

$$\leq h(2\mathbb{E}[|Y|] + c) + \tau \left(\frac{\log n}{n} \right)^{1/2}.$$

This is the desired result. $\qquad\square$

The following Corollary 1 is a consequence of Lemma 7.

Corollary 1. *Assume that $h = o((\log n/n)^{1/2})$. Then, for arbitrary $\tau < \kappa$, there exists some positive integer n_0, that does not depend of u, such that for any $n \geq n_0$, on the event $B_{c,h,\tau}(u)$ it holds that*

$$\forall v \in [0, u] : |\widehat{\varphi}_n(v) - \varphi(v)| \leq \kappa (\log n/n)^{\frac{1}{2}}.$$

Note that this integer n_0 is monotonically increasing with respect to τ and $\mathbb{E}[|Y|]$ and monotonically decreasing with respect to κ. Finally, for $u > 0$, define the event

$$B_\kappa(u)^c := \left\{ \exists v \in [0, u] : |\widehat{\varphi}(v) - \varphi(v)| > \kappa (\log n/n)^{1/2} \right\}.$$

Lemma 8. *Let $\kappa > \sqrt{1 + \tfrac{2}{K}}$. There exists some $C > 0$ which is monotonically increasing with respect to $\mathbb{E}[Y^2]$ such that for any $u \geq 1$ and $n \in \mathbb{N}$,*

$$\mathbb{P}(B_\kappa(u)^c) \leq C u n^{-\frac{1}{K}}.$$

Proof. Fix c and $\kappa > \tau$, assume that $h = o((n/\log n)^{-1/2}) > n^{-1/2}$. By Corollary 1, there exists, for any $\tau < \kappa$, some positive integer n_0, depending on $\mathbb{E}[|Y|]$ and τ such that for arbitrary $n \geq n_0$ we have $B_{c,h,\tau}(u) \subset B_\kappa(u)$. It follows that $\mathbb{P}(B_\kappa(u)^c) \leq \mathbb{P}(B_{c,h,\tau}(u)^c)$. The definition of $B_{c,h,\tau}(u)$ and Lemma 6 give, $\forall n \geq n_0$,

$$\mathbb{P}(B_{c,h,\tau}(u)^c) \leq \mathbb{P}(A_c^c) + 2\lceil u/h \rceil n^{-\tau^2/2} \leq \frac{1}{c^2 n} \mathbb{E}[Y^2] + 2\lceil u/h \rceil n^{-\tau^2/2}.$$

With the choice $\tau = \sqrt{1 + \tfrac{2}{K}}$, the statement of the Lemma follows. $\qquad\square$

14

6.2 Proof of Theorem 1

We have the decomposition

$$\|\widehat{f}_m - f\|^2 = \|f_m - f\|^2 + \|\widehat{f}_m - f_m\|^2 = \|f_m - f\|^2 + \frac{1}{2\pi}\int_{-m}^{m}|\widehat{\varphi}_X(u) - \varphi_X(u)|^2\,\mathrm{d}u.$$

The estimator $\widehat{\varphi}_X$ can be rewritten as follows

$$\widehat{\varphi}_X(u) = e^{\frac{1}{K}\widehat{\psi}(u)} = e^{\frac{1}{K}\mathrm{Re}\,\widehat{\psi}(u)}e^{\frac{1}{K}i\mathrm{Im}\,\widehat{\psi}(u)} = |\widehat{\varphi}(u)|^{\frac{1}{K}}e^{\frac{1}{K}i\mathrm{Im}\,\widehat{\psi}(u)},$$

where Re and Im denote the real and imaginary part of a complex number. For nonnegative real numbers, the K-th root is unique so that $|\varphi(u)|^{1/K} = |\varphi_X(u)|$. It follows that

$$\begin{aligned}|\widehat{\varphi}_X(u) - \varphi_X(u)|^2 =&\left||\widehat{\varphi}(u)|^{\frac{1}{K}}e^{\frac{i}{K}\mathrm{Im}\,\widehat{\psi}(u)} - |\varphi(u)|^{\frac{1}{K}}e^{\frac{i}{K}\mathrm{Im}\,\psi(u)}\right|^2\\ \leq&2\left||\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}}\right|^2 + 2|\varphi(u)|^{\frac{2}{K}}\left|e^{\frac{i}{K}\mathrm{Im}\,\widehat{\psi}(u)} - e^{\frac{i}{K}\mathrm{Im}\,\psi(u)}\right|^2.\end{aligned} \tag{6.9}$$

Consider the first term of (6.9). Using that $x \to x^{1/K}$ is Hölder continuous for $x > 0$ and the triangle inequality, we derive

$$\left||\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}}\right|^2 \leq \left|\widehat{\varphi}(u) - \varphi(u)\right|^{2/K}.$$

Integrating the former inequality in u, taking expectation and applying the Jensen inequality lead to

$$\mathbb{E}\left[\int_{-m}^{m}\left||\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}}\right|^2\,\mathrm{d}u\right] \leq \int_{-m}^{m}\left(\mathbb{E}[|\widehat{\varphi}(u) - \varphi(u)|^2]\right)^{1/K}\,\mathrm{d}u \leq mn^{-\frac{1}{K}}. \tag{6.10}$$

In the skewed case, we need to consider the second term of (6.9) that is nonzero. We derive that

$$\left|e^{\frac{i\mathrm{Im}\,\widehat{\psi}(u)}{K}} - e^{\frac{i\mathrm{Im}\,\psi(u)}{K}}\right|^2 \leq \frac{1}{K^2}\left|\mathrm{Im}\,\widehat{\psi}(u) - \mathrm{Im}\,\psi(u)\right|^2 \leq \frac{1}{K^2}\left|\int_0^u \frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)}\,\mathrm{d}v\right|^2.$$

Now, for arbitrary $v \in \mathbb{R}$,

$$\begin{aligned}\frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)} =&\frac{\varphi(v)\widehat{\varphi}'(v) - \widehat{\varphi}(v)\varphi'(v)}{\varphi(v)\widehat{\varphi}(v)} = \frac{\varphi(v)(\widehat{\varphi}'(v) - \varphi'(v)) - \varphi'(v)(\widehat{\varphi}(v) - \varphi(v))}{\varphi(v)\widehat{\varphi}(v)}\\ =&\frac{(\widehat{\varphi}'(v) - \varphi'(v))/\varphi(v) - (\varphi'(v)/\varphi(v)^2)(\widehat{\varphi}(v) - \varphi(v))}{\widehat{\varphi}(v)/\varphi(v)}\\ =&\frac{\left(-\frac{(\widehat{\varphi}(v) - \varphi(v))}{\varphi(v)}\right)'}{\left(1 - \frac{(\widehat{\varphi}(v) - \varphi(v))}{\varphi(v)}\right)}.\end{aligned} \tag{6.11}$$

First, we consider the case where m lies in $[0, u_n^{(\gamma,\epsilon)}]$, where $u_n^{(\gamma,\epsilon)}$ is defined in (2.4). Set $\kappa = (1 + \varepsilon)\gamma$, by Corollary 1 and by the definition of the event $B_\kappa(m)$ and of $u_n^{(\gamma,\epsilon)}$, we find that on $B_\kappa(m)$,

$$\forall u \in [-m, m] : |\widehat{\varphi}(u) - \varphi(u)| \leq \kappa\sqrt{\log n}n^{-\frac{1}{2}} \leq 1/(1 + \varepsilon)|\varphi(u)|. \tag{6.12}$$

15

Then, on the event $B_\kappa(m)$, a Neumann series expansion, along with formula (6.11) and formula (6.12) gives for $v \in [-m, m]$,

$$\frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)} = -\sum_{\ell=0}^{\infty} \Big(\frac{\widehat{\varphi}(v) - \varphi(v)}{\varphi(v)}\Big)' \Big(\frac{\widehat{\varphi}(v) - \varphi(v)}{\varphi(v)}\Big)^\ell.$$

The following representation holds

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

Moreover, we have $\widehat{\varphi}(0) - \varphi(0) = 0$ and get

$$\mathbb{1}_{B_\kappa(m)}\Big|\int_0^u \Big(\frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)}\Big)\,\mathrm{d}v\Big| \le \sum_{\ell=0}^{\infty} \frac{1}{\ell+1}\frac{|\widehat{\varphi}(u) - \varphi(u)|^{\ell+1}}{|\varphi(u)|^{\ell+1}}$$

$$\le \frac{|\widehat{\varphi}(u) - \varphi(u)|}{|\varphi(u)|}\sum_{\ell=0}^{\infty}\frac{(1+\varepsilon)^{-\ell}}{\ell+1} = (1+\varepsilon)\log(1+1/\varepsilon)\frac{|\widehat{\varphi}(u) - \varphi(u)|}{|\varphi(u)|}.$$

Gathering all the terms together, we have shown that for $m \in [0, u_n^{(\gamma,\epsilon)}]$,

$$\mathbb{E}\Big[\mathbb{1}_{B_\kappa(m)}\int_{-m}^m |\varphi_X(u)|^2\big|e^{\frac{i}{K}\mathrm{Im}\,\widehat{\psi}(u)} - e^{\frac{i}{K}\psi(u)}\big|^2\,\mathrm{d}u\Big]$$

$$\le \frac{c(\varepsilon)}{K^2}\int_{-m}^m |\varphi_X(u)|^2\frac{\mathbb{E}\big[|\widehat{\varphi}(u) - \varphi(u)|^2\big]}{|\varphi(u)|^2}\,\mathrm{d}u \le \frac{c(\varepsilon)}{K^2}\int_{-m}^m \frac{n^{-1}}{|\varphi_X(u)|^{2(K-1)}}\,\mathrm{d}u, \qquad (6.13)$$

where $c(\varepsilon) := (1+\varepsilon)\log(1+1/\varepsilon)$. On $B_\kappa(m)^c$, we use the majorant $|e^{i\mathrm{Im}\,\psi/K} - e^{i\mathrm{Im}\,\widehat{\Psi}/K}| \le 2$ and Lemma 8

$$\mathbb{E}\Big[\mathbb{1}_{B_\kappa(m)^c}\int_{-m}^m |\varphi_X(u)|^2\big|e^{\frac{i}{K}\mathrm{Im}\,\widehat{\psi}(u)} - e^{\frac{i}{K}\psi(u)}\big|^2\,\mathrm{d}u\Big] \le 4\mathbb{P}(B_\kappa(m)^c)\int_{-m}^m |\varphi_X(u)|^2\,\mathrm{d}u$$

$$\le 4\|\varphi_X\|^2 mn^{-\frac{1}{K}}. \qquad (6.14)$$

Secondly, we consider the case $m \ge u_n^{(\gamma,\epsilon)}$. The series of inequalities

$$|\widehat{\varphi}_X(u) - \varphi_X(u)| \le 2|\varphi_X(u)| + \big|\,|\widehat{\varphi}_X(u)| - |\varphi_X(u)|\,\big|$$

$$= 2|\varphi_X(u)| + \big|\,|\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}}\,\big| \le 2|\varphi_X(u)| + |\widehat{\varphi}(u) - \varphi(u)|^{\frac{1}{K}}, \qquad (6.15)$$

combined with the Jensen inequality, implies that

$$\int_{|u|\in[u_n^{(\gamma,\epsilon)},m]} \mathbb{E}\Big[|\widehat{\varphi}_X(u) - \varphi_X(u)|^2\Big]\,\mathrm{d}u \le 4\int_{|u|\in[u_n^{(\gamma,\epsilon)},m]} |\varphi_X(u)|^2\,\mathrm{d}u + 4\int_{|u|\in[u_n^{(\gamma,\epsilon)},m]} \mathbb{E}\Big[|\widehat{\varphi}(u) - \varphi(u)|^2\Big]^{\frac{1}{K}}\,\mathrm{d}u$$

$$\le 4\int_{|u|\in[u_n^{(\gamma,\epsilon)},m]} |\varphi_X(u)|^2\,\mathrm{d}u + 8n^{-\frac{1}{K}}m. \qquad (6.16)$$

Gathering equations (6.10), (6.13), (6.14) and (6.16) and defining $C_1 = c(\varepsilon)$ and $C_2 = (9 + 4\|\varphi_X\|^2)$, completes the proof of of Theorem 1. $\qquad\square$

16

6.3   Rates of convergence: Proof of Proposition 1.

Proposition 1 is a direct consequence of Theorem 1. To establish the result we bound independently each term of the right hand side of (2.5).

(i) Fix β, b, C, C', γ and ε. There exist a constant C_1 which depends on γ, ε and decreases with respect to C and a constant C_2 depending on a and decreasing with respect to C' such that for any $f \in \mathcal{F}(\beta, b, C, C', C_X, C_f)$ we have

$$u_n^{(\gamma, \varepsilon)} \geq C_1 (n / \log n)^{\frac{1}{2bK}} \quad \text{and} \quad u_n^{(a)} \leq C_2 n^{\frac{1}{2\beta K}}.$$

Using the fact that $m^* \leq u_n^{(a)}$ and the definition of $u_n^{(a)}$, we can estimate the second term in (2.5) as follows

$$\frac{1}{n} \int_{-m^*}^{m^*} \frac{1}{|\varphi_X(u)|^{2(K-1)}} \, du \leq \frac{2 u_n^{(a)}}{n} |\varphi_X(u_n^{(a)})|^{-2(K-1)} \leq 2 C_2 n^{\frac{1}{2\beta K} - 1} a n^{\frac{2(K-1)}{2K}}$$

$$= 2 a C_2 n^{-\frac{2\beta - 1}{2\beta K}} \leq 2 a C_2 n^{-\frac{2\beta - 1}{2bK}}.$$

The first and fourth term appearing in Theorem 1 can be controlled using that

$$\int_{\{|u| \geq u_n^{(\gamma, \varepsilon)}\}} |\varphi_X(u)|^2 \, du \leq \frac{2C'}{2\beta - 1} (u_n^{(\gamma, \varepsilon)})^{-(2\beta - 1)} \leq \frac{2C'}{2\beta - 1} \left(C_1 \frac{n}{\log n} \right)^{-\frac{2\beta - 1}{2bK}}.$$

Finally, the third term in (2.5) is bounded by

$$n^{-\frac{1}{K}} m^* \leq C_2 n^{-\frac{2\beta - 1}{2\beta K}} \leq C_2 n^{-\frac{2\beta - 1}{2bK}}.$$

We have thus shown that for a positive constant $\mathcal{C}n$ which depends on the choice of κ and ε, which increases with respect to C, C_X and C_f and which decreases with respect to β and C', we have that

$$\mathbb{E}[\|f - \widehat{f}_{m^*}\|^2] \leq \mathcal{C} n^{-\frac{2\beta - 1}{2bK}}.$$

(ii) Fix $\beta, \rho, b, c, C, C', \gamma$ and ε. Tere exists a constant C_1 depending on γ, ε and decreasing with C, c, b and a constant C_2 depending on a decreasing with respect to C' such that for any $f \in \mathcal{H}(\beta, \rho, b, c, C, C', C_X, C_f)$, we have

$$u_n^{(\gamma, \varepsilon)} \geq C_1 (\log n)^{\frac{1}{\rho}} \quad \text{and} \quad u_n^{(a)} \leq C_2 n^{\frac{1}{2\beta K}}.$$

Again, we may use the definition of m^* and the estimate for $u_n^{(a)}$ to derive that

$$\frac{1}{n} \int_{-m^*}^{m^*} \frac{1}{|\varphi_X(u)|^{2(K-1)}} \, du \leq 2 a C_2 n^{-\frac{2\beta - 1}{2\beta K}} = o\left((\log n)^{-\frac{2\beta - 1}{\rho}} \right).$$

Moreover, we have the upper bounds

$$\int_{\{|u| \geq u_n^{(\gamma, \varepsilon)}\}} |\varphi_X(u)|^2 \, du \leq \frac{2C'}{2\beta - 1} (C_1 \log n)^{-\frac{2\beta - 1}{\rho}}$$

17

and

$$m^* n^{-\frac{1}{K}} \le C_2 n^{-\frac{2\beta-1}{2\beta K}} = o\big((\log n)^{-\frac{2\beta-1}{\rho}}\big).$$

Putting the above together, we find that

$$\mathbb{E}[\|f - f_{m^*}\|^2] \le \mathcal{C}(\log n)^{-\frac{2\beta-1}{\rho}},$$

for some $\mathcal{C} \ge 0$ which is increasing with respect to C, c, b, C_X, C_f and decreasing with respect to β and C'.

(iii) Fix ρ, c, C, C', γ and ε. There exists a constant C_1 depending on γ and ε which is decreasing with respect to c and C and a constant C_2 depending on γ and ε which is decreasing with respect to c and C' such that for any $f \in \mathcal{G}(\rho, c, C, C', C_X, C_f)$, we have

$$u_n^{(\gamma,\varepsilon)} \ge C_1 (\log n)^{\frac{1}{\rho}} \quad \text{and} \quad u_n^{(a)} \le C_2 (\log n)^{\frac{1}{\rho}}.$$

In analogy with (i) and (ii), we have

$$\frac{1}{n} \int_{-m^*}^{m^*} \frac{1}{|\varphi_X(u)|^{2(K-1)}} \, du \le 2a C_2 n^{-\frac{1}{K}} (\log n)^{\frac{1}{\rho}}.$$

On the other hand, for some constant d depending on c, C and C', we derive from (2.8) that

$$\int_{\{|u| \ge u_n^{(\gamma,\varepsilon)}\}} |\varphi_X(u)|^2 \, du \le \frac{2C'}{c} \exp(-2c(u_n^{(\gamma,\varepsilon)})^\rho) \le d\big(u_n^{(\gamma,\varepsilon)}\big)^{1-\rho} \big|\varphi_X\big(u_n^{(\gamma,\varepsilon)}\big)\big|^2$$

$$\le dC_2 (\log n)^{\frac{1}{\rho}-1} \Big(\frac{n}{\log n}\Big)^{-\frac{1}{K}} \le dC_2 n^{-\frac{1}{K}} (\log n)^{\frac{1}{\rho}}.$$

We also get that

$$n^{-\frac{1}{K}} m^* \le C_2 n^{-\frac{1}{K}} (\log n)^{\frac{1}{\rho}}.$$

This implies for a constant $\mathcal{C}$ which depends on C, C' and c,

$$\mathbb{E}[\|f - f_{m^*}\|^2] \le \mathcal{C}(\log n)^{-\frac{2\beta-1}{\rho}},$$

which completes the proof of Proposition 1. $\qquad\square$

Remark. The dependence of $\mathcal{C}$ is not easy to express in a closed form. However, $\mathcal{C}$ is continuous in the constants involved.

6.4 Lower bound: Proof of Theorem 2

Following Theorem 2.7 in Tsybakov (2009), the lower bound is established by considering a decision problem between an increasing number of competing densities $f_0, \ldots, f_{M_n}$ contained in $\mathcal{F}(\beta, b, C, C', C_X, C_f)$, for some $M_n \ge 1$, such that the following two conditions are satisfied

(i) $\|f_j - f_k\|^2 \geq dn^{-\frac{2\beta-1}{2bK}}$, $\forall 0 \leq j < k \leq M_n$ for some positive constant d,

(ii) $\frac{1}{M_n} \sum_{j=1}^{M_n} \mathrm{KL}(\mathbb{P}_{j,K}^{\otimes n}, \mathbb{P}_{0,K}^{\otimes n}) \leq \alpha \log M_n$, $\forall 0 \leq j \leq M_n$ and for some $0 < \alpha < 1/8$,

where $\mathbb{P}_{j,K}$ denotes the probability measure corresponding to the density f_j^{*K} and KL is the Kullback-Leibler divergence. In the sequel, when there is no risk of confusion, $\mathcal{F}$ stands for the class $\mathcal{F}(\beta, b, C, C', C_X, C_f)$.

We start by constructing a density contained in $\mathcal{F}$, with characteristic function φ, such that

$$\liminf_{|u| \to \infty} |\varphi(u)||u|^b < \infty \quad \text{and} \quad \limsup_{|u| \to \infty} |\varphi(u)||u|^\beta > 0. \tag{6.17}$$

We start by considering the density

$$g_0(x) = \frac{3\sin^4(x)}{2\pi x^4}.$$

The characteristic function $\varphi_{g,0}$ of g_0 is, up to a multiplicative factor $3/2$, the 4-fold auto-convolution of the distribution $\mathcal{U}([-1, 1])$ and hence supported on $[-4, 4]$.

To ensure the second part of (6.17), we define for some positive constant c,

$$g(x) = g_0(x) + c \sum_{k=1}^{\infty} (1 + (8k))^{-\beta} \cos(12kx) g_0(x).$$

Using the boundedness of the cosine-function, along with $\beta > 1$ and an appropriate choice of c, g is nonnegative. Moreover, since g integrates to 1, g is a density. The multiplication with $\cos(12kx)$ corresponds to a left and right shift in the Fourier domain, so the characteristic function of g is

$$\varphi_g(u) = \varphi_{g,0}(u) + \frac{c}{2} \sum_{k=1}^{\infty} (1 + (8k))^{-\beta} (\varphi_{g,0}(u + 12k) + \varphi_{g,0}(u - 12k)). \tag{6.18}$$

For g constructed as above, we can a priori not guarantee that (2.7) is valid. However, this can be ensured with an appropriate rescaling.

So far, g is not contained in the class $\mathcal{F}$, since the absolute value of φ_g is not bounded from below, it does even have zeros. To ensure the lower bound on the characteristic function under consideration, we proceed as follows. Let h be a symmetric density which decays, in the time domain, as x^{-4} and whose characteristic function decays as $|u|^{-b}$. For example, h can be selected as the mixture of a density of the form $c'(1 + x^2)^{-2}$ and of a symmetric bilateral $\Gamma(b/2, \lambda)$ density.

Set

$$f_0(x) = c_1 g(x) + c_2 h(x).$$

For a positive constants c_1 to be chosen and $c_2 = 1 - c_1$. Finally, by construction f_0 belongs to the class $\mathcal{F}$.

19

A collection of competing densities $f_0, \ldots, f_{M_n}$ is obtained by considering perturbed versions of the characteristic function φ_0 of f_0. Let ψ be a complex-valued function satisfying the following properties:

- ψ is supported on $[-8, 8]$.

- $\psi(u) = 1 - e^{i2\pi/K}$, $\forall u \in [-4, 4]$.

- ψ has three continuous (complex) derivatives.

- $|1 - \psi(u)| \geq 1/2$ $\forall u \in [-8, 8]$.

- $|\psi(u)| \leq 2$ $\forall u \in [-8, 8]$.

For $n \in \mathbb{N}$, a family of perturbations is defined as follows. Let $L_n = 12k_n$, with some $k_n \in \mathbb{N}$ to be specified. Let m_n be a positive integer to be chosen. For $1 \leq m \leq m_n$ define

$$\psi_m(u) = \psi(u - L_n - 12m) + \overline{\psi}(u + L_n + 12m).$$

Given any subset $\mathcal{M} \subseteq \{1, \ldots, m_n\}$, a perturbed version of φ_0 is defined by

$$\varphi_{\mathcal{M}}(u) := \varphi_0(u) - \sum_{m \in \mathcal{M}} \psi_m(u) \varphi_0(u).$$

In the sequel, define the sets

$$I_{n,m} := [-L_n - 12m - 4, -L_n - 12m + 4] \cup [L_n + 12m - 4, L_n + 12m + 4],$$

and

$$J_{n,m} := \big([-L_n - 12m - 8, -L_n - 12m + 8] \cup [L_n + 12m - 8, L_n + 12m + 8]\big) \setminus I_{n,m},$$

which are disjoints for distinct values of m. Note that the following holds, for all $m \in \mathcal{M}$. The function ψ_m is compactly supported on the interval $I_{n,m} \cup J_{n,m}$ and if $u \in I_{n,m}$ we have $\psi_m(u) = (1 - e^{i2\pi/K})$.

Now, if we consider $\varphi_g(u) - \sum_{m \in \mathcal{M}} \psi_m(u) \varphi_g(u)$, by construction, this function is equal to the function $\varphi_g e^{i2\pi/K}$ on any interval of the form $I_{n,m}$, $m \in \mathcal{M}$ and with φ_g elsewhere since on the interval $J_{n,m}$, the function φ_g is zero. Using the same arguments as for formula (6.18), we find that the inverse Fourier transform of this quantity is non negative and hence is a density. Next, consider $\varphi_h(u) - \sum_{m \in \mathcal{M}} \psi_m(u) \varphi_h(u)$. By construction of ψ, the inverse Fourier transform of ψ_m decays faster than h and the inverse Fourier transform of $\psi_m \varphi_h$ is bounded by $cL_m^{-b}|\mathcal{F}^{-1}\psi_m|$. It follows that the inverse Fourier transform of $\varphi_h(u) - \sum_{m \in \mathcal{M}} \psi_m(u) \varphi_h(u)$ is non-negative, provided that $m_n \leq L_m/c$. Finally, by the definition of ψ, $\varphi_{\mathcal{M}}$ has the same asymptotic behavior as φ_0 for any subset $\mathcal{M}$. Finally, $\varphi_{\mathcal{M}}$ belongs to $\mathcal{F}$.

We are now ready to check that (i) holds by bounding from below the L_2-distance between the competing densities. By definition of the ψ_m, given two sets of indices $\mathcal{M}_1, \mathcal{M}_2 \subseteq \{m_1, \ldots, m_n\}$, we have

$$
\begin{aligned}
\big|\varphi_{\mathcal{M}_1}(u)|_{I_{n,m}} - \varphi_{\mathcal{M}_2}(u)|_{I_{n,m}}\big| &= \mathbb{1}_{\mathcal{M}_1 \triangle \mathcal{M}_2}(m)|1 - e^{i2\pi/K}|\big|\varphi_0(u)|_{I_{n,m}}\big| \\
&= \mathbb{1}_{\mathcal{M}_1 \triangle \mathcal{M}_2}(m)2|\sin(\pi/K)|\big|\varphi_0(u)|_{I_{n,m}}\big|,
\end{aligned}
$$

20

with $\triangle$ denoting the symmetric difference. Using this, along with formula (6.18), we derive that there exists a positive constant C such that

$$\|\varphi_{\mathcal{M}_1} - \varphi_{\mathcal{M}_2}\|^2 \geq C|\mathcal{M}_1 \triangle \mathcal{M}_2|(1 + |L_m + 12m_n|)^{-2\beta} \int_{-4}^{4} |\varphi_{g,0}(u)|^2 \, du. \tag{6.19}$$

The Varshamov-Gilbert bound (see, e.g. Lemma 2.9 in Tsybakov (2009)) guarantees that for $m_n \geq 8$, there exist a family $\{\mathcal{M}_{n,k} : 0 \leq k \leq M_n\}$ of subsets of $\{1, \ldots, m_n\}$ such that

- $\mathcal{M}_{n,0} = \emptyset$,

- $M_n \geq 2^{m_n/8}$ and

- $|\mathcal{M}_{n,k} \triangle \mathcal{M}_{n,j}| \geq m_n/8, \ \forall \, 0 \leq j < k \leq M_n$.

Combining this with formula (6.19), we find that for $n \in \mathbb{N}$ large enough, there exist $M_n = 2^{m_n/8}$ densities $f_0, \ldots, f_{M_n}$ belonging to $\mathcal{F}$ and for which

$$\|f_j - f_k\|^2 \geq Cm_n m_n^{-2\beta}, \ j \neq k$$

holds, provided that $L_n \asymp m_n$.

Let us now bound the Kullback-Leibler divergence to establish (ii). Using that the Kullback-Leibler divergence is bounded by the χ^2-distance, as well as the fact that

$$\mathrm{KL}(\mathbb{P}^{\otimes n}, \mathbb{Q}^{\otimes n}) = n \, \mathrm{KL}(\mathbb{P}, \mathbb{Q}),$$

it is enough to consider the χ^2-distance between the competing densities $f_j^{*K}, \ j = 0, \ldots, M_n$,

$$\chi^2(\mathbb{P}_{j,K}, \mathbb{P}_{0,K}) = \int_{-\infty}^{\infty} \frac{(f_j^{*K}(x) - f_0^{*K}(x))^2}{f_0^{*K}(x)} \, dx.$$

Since f_0^{*K} and f_0 decay at the same rate x^{-4}, we have a some positive constant C,

$$\chi^2(\mathbb{P}_{j,K}, \mathbb{P}_{0,K}) \leq C \int_{-\infty}^{\infty} (1 + x^4)(f_j^{*K}(x) - f_0^{*K}(x))^2 \, dx$$

$$= C \int_{-\infty}^{\infty} (f_j^{*K}(x) - f_0^{*K}(x))^2 \, dx + C \int_{-\infty}^{\infty} (x^2 f_j^{*K}(x) - x^2 f_0^{*K}(x))^2 \, dx.$$

The multiplication with x^k corresponds, up to the factor i^{-k}, to taking the k-th derivative in the Fourier domain. The Plancherel formula gives

$$\chi^2(\mathbb{P}_{j,K}, \mathbb{P}_{0,K}) \leq \frac{C}{2\pi} \int_{-\infty}^{\infty} |\varphi_j^K(u) - \varphi_0^K(u)|^2 \, du + \frac{C}{2\pi} \int_{-\infty}^{\infty} |(\varphi_j^K)''(u) - (\varphi_0^K)''(u)|^2 \, du.$$

By construction, on the intervals $I_{n,k}, \ k \in \mathcal{M}_n$, φ_0 is strictly positive, we have

$$\varphi_j(u) = \varphi_0(u)e^{i2\pi/K} = |\varphi_0(u)|e^{i2\pi/K}.$$

21

Consequently, φ_j^K and φ_0^K are equal on $I_{n,k}$. For $u \in J_{n,m}$, $m \in \mathcal{M}_n$,

$$(\varphi_j^K(u) - \varphi_0^K(u)) = \varphi_0(x)^K (1 - (1 - \psi_m)^K(u)).$$

Now, on the intervals $J_{n,m}$, φ_0 agrees with φ_h. As h is the mixture of a density in $C(1 + x^2/c^2)^{-2}$ and a bilateral gamma density, we can derive from the analysis given in Küchler and Tappe (2008) that $|(\varphi_0^K)^{(\ell)}(u)| \lesssim |u|^{-Kb-\ell}$, $\ell = 0,1,2$. Along with the boundedness of the derivatives of $\psi(u)$, this implies that for $u \in J_{n,m}$,

$$|\varphi_j^K(u) - \varphi_0^K| + |(\varphi_j^K)''(u) - (\varphi_0^K)''(u)|^2 \lesssim |u|^{-2bK}.$$

Finally, $\varphi_j^K = \varphi_0^K$ on the complement of the set $\cup_{m \in \mathcal{M}_n}(I_{n,m} \cup J_{n,m})$. It follows that

$$\chi^2(\mathbb{P}_{j,K}, \mathbb{P}_{0,K}) \lesssim m_n L_n^{-2bK},$$

and

$$\frac{1}{M_n} \sum_{k=1}^{M_n} \mathrm{KL}(\mathbb{P}_{j,K}^{\otimes n}, \mathbb{P}_{0,K}^{\otimes n}) \lesssim n m_n L_n^{-2bK} \leq (1/8 \log M_n) n L_n^{-2bK}.$$

The choice $L_n \asymp m_n \asymp n^{\frac{1}{2bK}}$ and Theorem 2.7 in Tsybakov (2009) imply the result. $\qquad\square$

6.5   Adaptive estimation: Proof of Theorem 3.

In the sequel, we write $\widehat{m}$ instead of $\widehat{m}_\eta$. Recall that, for some $\eta > 1$,

$$\widehat{m} = \min\Big\{ \min\{u > 0 : |\widehat{\varphi}(u)| = (nK)^{-\frac{1}{2}} + \sqrt{\frac{\eta}{K}\log n}\, n^{-\frac{1}{2}}\}, n^{\frac{1}{K}}\Big\}.$$

In what follows, $\mathcal{E}$ stands for one of the function classes, namely, $\mathcal{F}(\beta, b, C, C', C_X, C_f)$, $\mathcal{H}(\beta, \rho, b, c, C, C', C_X, C_f)$ or $\mathcal{G}(\rho, c, C, C', C_X, C_f)$. Depending on the class considered, let $r_n = (n/\log n)^{\frac{2\beta-1}{2bK}}$, $r_n = (\log n)^{\frac{2\beta-1}{\rho}}$ or $r_n = (\log n)^{\frac{1}{\rho}} n^{-\frac{1}{K}}$.
Let f be any density in $\mathcal{E}$ and define

$$m_0 := \min\{u > 0 : |\varphi(u)| = (2\sqrt{\eta/K} + \kappa)(\log(n)/n)^{1/2}\}$$

and

$$m_1 := \min\{u > 0 : |\varphi(u)| = (nK)^{-\frac{1}{2}}\}.$$

Firstly, using the definition of m_0, along with Lemma 8 and the triangle inequality, we find that

$$\mathbb{P}\big(\widehat{m} < m_0\big) \leq \mathbb{P}\big(\exists u \in [0, m_0] : |\widehat{\varphi}(u) - \varphi(u)| \geq \kappa\sqrt{\log n}\, n^{-\frac{1}{2}}\big) \leq C m_0 n^{-\frac{1}{K}}, \qquad (6.20)$$

with some constant C depending on $\mathbb{E}[X^2] < \overline{C}_X$.
Secondly, by the Hoeffding inequality, the triangle inequality and by the continuity of the empirical characteristic function,

$$\mathbb{P}\big(\widehat{m} > m_1\big) \leq \mathbb{P}\big(|\widehat{\varphi}(m_1)| > (nK)^{-\frac{1}{2}} + \sqrt{\eta\log(n)/K}\, n^{-\frac{1}{2}}\big)$$
$$\leq \mathbb{P}\big(|\varphi(m_1) - \widehat{\varphi}(m_1)| > \sqrt{\eta\log(n)/K}\, n^{-\frac{1}{2}}\big) \leq n^{-\frac{\eta}{K}}. \qquad (6.21)$$

22

The proof of Theorem 1 implies that

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\{\widehat{m} \in [m_0, m_1]\}}]$$

$$\leq \|f - f_{m_0}\|^2 + \frac{C_1}{nK^2} \int\limits_{-m_1}^{m_1} \frac{1}{|\varphi_X(u)|^{2(K-1)}} \, \mathrm{d}u + C_2 m_1 n^{-\frac{1}{K}} + \frac{2}{\pi} \int\limits_{[u_n^{(\kappa,\epsilon)}, m_1]} |\varphi_X(u)|^2 \, \mathrm{d}u, \quad (6.22)$$

with a constant C_1 depending on the definition of ε and with C_2 depending on $\|f\| < \overline{C}_f$ and on $\mathbb{E}[X^2] < \overline{C}_X$. We can now use the same arguments as in the proof of Proposition 2 to derive from (6.22) that

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\{\widehat{m} \in [m_0, m_1]\}}] \leq \mathcal{C}_1 r_n \qquad (6.23)$$

for a positive constant $\mathcal{C}_1$ depending on I and on the choice of the parameters.
Consider now the exceptional set $\widehat{m} \notin [m_0, m_1]$. Applying the Plancherel formula and arguing along the same lines as in the proof of Theorem 1, we derive that

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\{\widehat{m} < m_0\}}] \leq \mathbb{P}(\widehat{m} < m_0)\|f\|^2 + \frac{1}{2\pi} \int_{-m_0}^{m_0} \mathbb{E}[|\widehat{\varphi}_X(u) - \varphi_X(u)|^2] \, \mathrm{d}u$$

$$\leq \mathbb{P}(\widehat{m} < m_0)\|f\|^2 + \frac{C_1}{nK^2} \int_{-m_0}^{m_0} \frac{1}{|\varphi_X(u)|^2} \, \mathrm{d}u + C_2 n^{-\frac{1}{K}} m_0 + \frac{2}{\pi} \int\limits_{[u_n^{(\gamma,\epsilon)}, m_0]} |\varphi_X(u)|^2 \, \mathrm{d}u. \qquad (6.24)$$

Finally, the fact that $\widehat{m} \leq n^{\frac{1}{K}}$, along with the Plancherel formula and formula (6.15) leads to

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\{\widehat{m} > m_1\}}]$$

$$\leq \|f - f_{m_1}\|^2 + 4\mathbb{P}(\widehat{m} > m_1) \int_{-n^{\frac{1}{K}}}^{n^{\frac{1}{K}}} |\varphi_X(u)|^2 \, \mathrm{d}u + \int_{-n^{\frac{1}{K}}}^{n^{\frac{1}{K}}} \mathbb{E}[\mathbb{1}_{\{\widehat{m} > m_1\}} |\widehat{\varphi}(u) - \varphi(u)|^{\frac{1}{K}}] \, \mathrm{d}u.$$

It follows, using successively the Hölder, the Jensen and the Rosenthal inequalities, that for some constant C and some C' depending on η,

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\{\widehat{m} > m_1\}}] \leq \|f - f_{m_1}\|^2 + Cn^{-\frac{1}{K}}\|\varphi_X\|^2 + C'\mathbb{P}(\widehat{m} > m_1)^{\frac{1}{\eta}} \int_{-n^{\frac{1}{K}}}^{n^{\frac{1}{K}}} n^{-\frac{1}{K}} \, \mathrm{d}u. \quad (6.25)$$

Combining (6.24) and (6.25) together with (6.20) and (6.21) and plugging in m_0 and m_1, we have shown that for a positive constant $\mathcal{C}_2$ depending again on I and the choice of ε, δ and η,

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbb{1}_{\widehat{m} \notin [m_0, m_1]}] \leq \mathcal{C}_2 r_n. \qquad (6.26)$$

Putting (6.23) and (6.26) together, we have shown that there exists a constant $\mathcal{C}$ depending on the choice of η, δ and ε and on I, such that for any $f \in \mathcal{E}$,

$$\frac{\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2]}{r_n} \leq \mathcal{C}.$$

Taking the suprema gives the statement of the theorem. $\qquad \square$

23

References

[1] Buchmann, B. and Grübel, R. (2003). Decompounding: an estimation problem for Poisson random sums. *The Annals of Statistics,* **31**, 1054–1074.

[2] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics*, 32(2):181–192.

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

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

[5] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* , 83(404):1184–1186.

[6] Comte, F. and Genon-Catalot, V. (2010). Nonparametric adaptive estimation for pure jump Lévy processes. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, **46**, 595–617.

[7] Comte, F., Duval, C., Genon-Catalot, V. and Kappus, J. (2015). Estimation of the jump size density in a mixed compound Poisson process. *Scandinavian Journal of Statistics*, **42**, 1023–1044.

[8] Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *Journal of Multivariate Analysis*, **140**, 21-46.

[9] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B* , 73:601–627.

[10] Comte, F., Rozenholc, Y. and Taupin M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simul.*, 77(11-12):977–1000.

[11] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2):665–685.

[12] Delaigle, A. and Zhou, W. (2015). Nonparametric and parametric estimators of prevalence from group testing data with aggregated covariates. *Journal of the American Statistical Association,*. To appear.

[13] Delattre, S., Hoffmann, M., Picard, D. and Vareschi, T. (2012). Blockwise SVD with error in the operator and application to blind deconvolution. *Electron. J. Stat.*, 6:2274–2308.

[14] Duval, C. (2014) When is it no longer possible to estimate a compound Poisson process? *Electronic Journal of Statistics* **8**, 274–301.

[15] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding, *Bernoulli*, **13**, 672–694.

[16] Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics* , 19(3):1257–1272.

[17] Johannes, J. (2009). Deconvolution with unknown error distribution. *The Annals of Statistics* , 37(5a):2301–2323.

[18] Johannes, J. and Schwarz, M. (2013). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli* , 19(5A):1576–1611.

[19] Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequancy. *Stochastic Processes and their Applications* **124**, 730–758.

[20] Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution, *Electronic Journal of Statistics*, **8**, 2879-2904.

[21] U. Küchler and S. Tappe. On the shapes of bilateral Gamma densities. *Statist. Probab. Lett.*, 78(15):2478–2484, 2008.

[22] Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *Comptes Rendus Mathématique*, **342**(11), 877-882.

[23] O. Linton and Y.-J. Whang. Nonparametric estimation with aggregated data. *Econometric Theory*, 18(2):420–468, 2002.

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

[25] Meister, A. (2007). Optimal convergence rates for density estimation from grouped data. *Statistics & probability letters*, **77**(11), 1091–1097.

[26] Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Lecture Notes in Statistics. Springer.

[27] Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* , 7(4):307–330.

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

[29] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics* , 27(6):2033– 2053.

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

[31] Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics and Probability Letters* , 9(3):229–235.

[32] Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics* , 21(2):169–184

[33] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.