



HAL
open science

Nonparametric adaptive estimation for grouped data

Céline Duval, Johanna Kappus

► **To cite this version:**

Céline Duval, Johanna Kappus. Nonparametric adaptive estimation for grouped data. 2015. hal-01245781v1

HAL Id: hal-01245781

<https://hal.science/hal-01245781v1>

Preprint submitted on 17 Dec 2015 (v1), last revised 3 Jun 2016 (v2)

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

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

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 Fourier transform of the observations. We propose a new strategy for the data driven choice of the cut-off parameter. It is proven optimal and a numerical study illustrates the performances of the method. Moreover, we discuss the fact that the definition of the estimator as well as the adaptive procedure apply in a wider context than the one considered here.

Keywords. Convolution. Inverse problem. Nonparametric adaptive estimation.

AMS Classification. 62G07, 62G20, 62G05.

1 Introduction

Let X be some real valued random variable which has a square integrable Lebesgue density f . Suppose that we aim at estimating f , but only have access to the aggregated data

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

where the $X_{j,k}$, $k = 1, \dots, K$; $j = 1, \dots, n$ are independent copies of X and K is a known positive integer. An example where related models are encountered in practical applications is in large screening studies where infectious disease are involved. As the cost of individual measurement might be important, individuals are grouped before being observed. Recently, Delaigle and Zhou (2015) studied the nonparametric estimation of the probability of contamination given a covariate X that may represent, for instance, a pollutant, when only grouped data Y (with our notations) are observed.

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. In the Fourier domain, 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. An estimator of f can hence be constructed, replacing φ_Y by its empirical counterpart. However, this approach suffers from

*¹MAP5, UMR CNRS 8145, Université Paris Descartes.

†²Institut für Mathematik, Universität Rostock

the fact that the definition of K -th root is ambiguous, unless φ_X is real and positive, i.e. f is symmetric. In this paper we provide a well posed definition of the K -th root, based on a suitable definition of the logarithm for characteristic function that is not necessarily real. We naturally end up with a procedure to estimate f .

Meister (2007) studied the case where f is symmetric and proposed a cross-validation bandwidth selector. However, theoretical optimality properties for the adaptive procedure are only shown under quite restrictive assumptions on the characteristic function. An estimator for the non-symmetric case is also discussed in the afore mentioned paper and a modified version of it is used by Delaigle and Zhou (2015). However, this estimator has the disadvantage that it is not given in a closed form, making its the practical calculation difficult to handle.

In the present paper, a fully constructive estimation procedure is introduced which covers both the non-symmetric and symmetric case. Moreover, we propose a new data driven selection strategy for the cut-off parameter and show that the resulting rates of convergence are adaptive minimax under very general assumptions. Our approach has the clear advantage of being computationally simple and extremely fast, in comparison to usual cross-validation techniques.

The interest of the present article is threefold. First, we offer a unified adaptive and optimal strategy to estimate the underlying distributional density when grouped data are available. The model considered here belongs to the broader class of statistical inverse problems and can also be compared to the problem of estimating the jump density from a random sum, see e.g. Buchmann and Grübel (2003) or van Es et al. (2007).

Second, our estimator is defined as the empirical counterpart of the logarithm of a characteristic function, for which an optimal upper bound is established. This estimator and the techniques we introduce to derive the associated upper bound apply beyond the scope of density estimation from grouped data.

Finally, as it is usually the case when an estimator is built on the characteristic function, an optimal cut-off needs to be adaptively selected from the observations. The data driven selection method we introduce leads to an optimal rate estimator. This procedure applies in a wider context than the one considered here and performs numerically fast.

The paper is organized as follows. Section 2 contains the construction of the estimator. Upper 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.

2 Risk bounds and rates of convergence

Some notations

We start by introducing some notations 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, $f^{*K} = f * \dots * f$. Given a random variable Z ,

$\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$ denotes the characteristic function. For Y_j defined as in formula (1), we will drop the subscript and write φ instead of φ_{Y_j} . For $f \in L^1(\mathbb{R})$, $\mathcal{F}f(u) = \int e^{iux} f(x) dx$ is understood to be the Fourier transform. Moreover, we denote by $\|\cdot\|$ the L^2 -norm of functions, $\|f\|^2 := \int |f(x)|^2 dx$. Given some function $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, we denote by f_m the uniquely defined function with Fourier transform $\mathcal{F}f_m = (\mathcal{F}f)\mathbb{1}_{[-m,m]}$.

2.1 Construction of the estimator

In the sequel, we will always work under the assumption

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

This assumption is standard in deconvolution problems and, for the present model, proven necessary in Meister (2007). Therein, a counterexample is provided when K is even and (A0) is not satisfied.

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,

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

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 non-zero imaginary part, the definition of $z \mapsto z^{1/K}$ is not unique. So one needs to specify which branch of the logarithm is considered in order to get a proper definition of the K -th root. 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 $\text{Log } \varphi$. Moreover, in the sequel, we let $1/K$ denote the *distinguished K -th root*, that is, $\varphi^{1/K} := \exp((\text{Log } \varphi)/K)$.

It is important to keep in mind that the distinguished logarithm does usually not agree with the main branch of the logarithm. Moreover, the definition of a distinguished logarithm and 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,*

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

The proof of Lemma 1 and Lemma 2 can be found, for example, in Chapter 7 in Sato (2005). 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 zero points and assume that φ is differentiable. Then, it follows that*

$$\text{Log}(\varphi(u)) = \int_0^u \frac{\varphi'(z)}{\varphi(z)} dz. \quad (2.2)$$

The preceding lemmas suggest to exploit formula (3) 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

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

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 ,

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

It is worth pointing out that for real valued and strictly positive characteristic functions, our estimator coincides with the estimator given in Meister (2007) since in this case, the distinguished K -th root equals the usual K -th root. Also our estimator resembles one of the estimator defined in Comte et al. (2015), which considers the problem of estimating the jump density for mixed compound Poisson processes. However, in Comte et al. (2015) the associated upper bound is non adaptive and sub optimal.

2.2 Non-asymptotic risk bounds

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

$$u_{\varepsilon, \gamma} := \min\{u \geq 0 : |\varphi(u)| = (1 + \varepsilon)\gamma(n/\log n)^{-1/2}\}. \quad (2.3)$$

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 $\gamma := \sqrt{\frac{1}{2} + \frac{1}{K} + \delta}$, for some $\delta > 0$. Then, there exist positive constants C_1 and C_2 depending on the definition of ε and the choice of δ , which are increasing with respect to $\mathbb{E}[X^2]$ and $\|f\|$, such that for any $m \geq 0$,*

$$\mathbb{E}[\|f - \hat{f}_m\|^2] \leq \|f - f_m\|^2 + \frac{C_1}{nK^2} \int_{-m}^m \frac{1}{|\varphi_X(u)|^{2(K-1)}} du + C_2 m n^{-\frac{1}{K}} + \frac{2}{\pi} \int_{|u| \in [u_{\varepsilon, \gamma}, m]} |\varphi_X(u)|^2 du,$$

with the convention that if $m \leq u_{\varepsilon, \gamma}$ then $[u_{\varepsilon, \gamma}, m] = \emptyset$.

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. Here, f 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)). It directly follows from the proof of Theorem 1, that in the symmetric case the second summand does not intervene. Finally, the fourth summand depends on the quantity $u_{\varepsilon, \gamma}$, that appears naturally. Indeed, in the definition of $\hat{\varphi}_X$, the quantity $\hat{\varphi}^{-1}$ intervene, which may get close to 0. We prove that on the interval $[-u_{\varepsilon, \gamma}, u_{\varepsilon, \gamma}]$, the estimator $\hat{\varphi}$ can be controlled uniformly.

Balancing the four terms, there will occur a logarithmic loss, in the non symmetric case, in comparison to a standard deconvolution problem. This phenomenon is a consequence of the fact that the estimation problem in the non symmetric case involves distinguished logarithm, which requires uniform rather than point wise control of the empirical characteristic function. Controlling suprema will typically lead to the loss of a logarithmic factor.

2.3 Rates of convergence

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

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

and, in addition

$$\mathbb{E}[X^2] \leq C_X \quad \text{and} \quad \|f\| \leq C_f. \quad (2.4)$$

For $\rho > 0$, denote by $\mathcal{F}(\beta, \rho, c, C, C', C_X, C_f)$ the class of densities for which (2.3) holds and, in addition,

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

The following result is an immediate consequence of Theorem 1.

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

$$u_a := \min\{u \geq 0 : |\varphi(u)| = a n^{-\frac{1}{2}}\}.$$

Fix $a > 0$ and γ as in Theorem 1. For every $n \in \mathbb{N}$, let $m^ \in [u_{\varepsilon, \gamma}, u_a]$. Then, we have*

$$(i) \quad \sup_{f \in \mathcal{F}(\beta, C, C', C_X, C_f)} \mathbb{E}[\|f - \hat{f}_{m^*}\|^2] = O\left((n/\log n)^{-\frac{(2\beta-1)}{2K\beta}}\right).$$

$$(ii) \quad \sup_{f \in \mathcal{F}(\beta, \rho, c, C, C', C_X, C_f)} \mathbb{E}_f[\|f - \hat{f}_{m^*}\|^2] = O\left(n^{-\frac{1}{K}}(\log n)^{\frac{1}{\rho}}\right).$$

It is straightforward to see that $u_{\varepsilon, \gamma} < u_a$ so that $[u_{\varepsilon, \gamma}, u_a]$ is non empty. The rates of convergence derived for our constructive estimator coincide with the rates found in Meister (2007), where there are proven to be minimax optimal. The rate results easily generalize to different decay scenarios of the characteristic function. More specifically it is possible to apply Theorem 1 to a class of densities $\mathcal{F}(h_1, h_2)$ such that, for some positive functions h_1 and h_2 , we have

$$\forall u \in \mathbb{R} : 0 < h_1(u) \leq |\varphi_X(u)| \leq h_2(u),$$

where $\frac{h_2(u)}{h_1(u)} \rightarrow \ell \in (0, \infty]$ as $|u| \rightarrow \infty$. The resulting rate will depend on how h_1 and h_2 converge to 0 at infinity.

3 Adaptive estimation

It is interesting to notice that the theoretical cut-off parameter m^* , which may vary in a certain interval and guarantees rate optimality of the estimator, has an empirically accessible counterpart. The cut-off m^* is defined in terms of the characteristic function: the estimator of $\hat{\varphi}_X$ 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.

This inspires to define the empirical cut-off in terms of the empirical characteristic function. Once $\hat{\varphi}$ is (up to an additional logarithmic term) below some critical threshold, $\hat{\varphi}_X$ is set to 0. Fix a constant $\eta > 1$, to be chosen, let us introduce the data driven version of m^*

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

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

Theorem 2. (i) Let constants $B, \bar{C}, \bar{C}', \bar{C}_X$ and \bar{C}_f be given. Define $\bar{I} := (1/2, B] \times (0, \bar{C}] \times (0, \bar{C}') \times (0, \bar{C}_X] \times (0, \bar{C}_f]$ Then there exists some positive real \mathcal{C} such that

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

(ii) Let constants $B, R, \bar{c}, \bar{C}, \bar{C}', \bar{C}_X$ and \bar{C}_f be given. Define $\bar{I} := (0, B] \times (0, R] \times (0, \bar{c}] \times (0, \bar{C}] \times (0, \bar{C}') \times (0, \bar{C}_X] \times (0, \bar{C}_f]$ Then there exists some positive real \mathcal{C} such that

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

In earlier research on the subject, the adaptive bandwidth selection has only been investigated for the particular case of symmetric densities and for polynomially decaying characteristic functions. Moreover, in comparison to the cross-validation strategy proposed in Meister (2007), our adaptive estimator has the strong advantage of being computationally extremely simple and fast, since it only requires the evaluation of the characteristic function.

Proof of Theorem 2 does not use the fact that we deal with grouped data. In fact, the bandwidth selector can straightforwardly be used for standard density estimation.

The adaptive choice \widehat{m}_η depends on the constant η that needs to be chosen. This is not unusual when it comes to adaptation that an hyper parameter needs to be calibrated. It is the case in the context of wavelet adaptive estimation or when the spectral cut-off is obtained by minimization of a penalty.

4 Discussion

The grouped data model can also be compared to the estimation of the jump density of a random sum, see e.g. Buchmann and Grübel (2003), van Es et al. (2007), Bøgsted and Pitts (2010) or Duval (2013a,2013b). In these models, the jump density can, up to some additive constant, be expressed in terms of the (distinguished) logarithm of the characteristic function of the distributional density.

It is worth noticing that if K increases, the rates of convergence deteriorate. If we let $K \rightarrow \infty$ then, (see Meister (2007) for the lower bound results)

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

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 , we have

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

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, 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 $\varphi_1 = \exp((\text{Log } \varphi_\Delta)/\Delta)$. The estimation procedure proposed in this paper carries over immediately, leading to the estimator

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

and the upper bound

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

Balancing the terms when φ_1 is polynomially decreasing leads to a rate in $(T/\log T)^{-\frac{2\beta-1}{2\Delta\beta}}$ 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}_η 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 on various examples. We also compare its performances with the cross-validation estimator proposed in Meister (2007), even though, no rigorous 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. Finally, we compare our procedure with an oracle estimator, more precisely the estimator \widehat{f}_{m^*} that corresponds to the oracle bandwidth

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

The L^2 risks are computed after 500 Monte-Carlo simulations of each estimator and averaging the results. They are denoted r , r_{cv} , and r_{or} , respectively.

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.
- (iii) Gamma distribution with parameters 6 and 3.
- (iv) Laplace function with mean 0.5 and scale parameter $\beta = 3$.

Moreover, we set $\eta = 1.1$ for the adaptive procedure. 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

Our procedure performs well in all examples. The risks decrease as n increases. As expected, 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

$$|\varphi(u) - \varphi(u + h)| = |\mathbb{E}[e^{iuZ} - e^{i(u+h)Z}]| \leq h\mathbb{E}[|Z|].$$

□

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 $h > 0$, 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 using that $x \rightarrow e^{ix}$ 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

$$\frac{h}{n} \sum_{j=1}^n |Y_j| = h(\mathbb{E}[|Y|] + \sum_{j=1}^n (|Y_j| - \mathbb{E}[|Y|])) \leq h(\mathbb{E}[|X|] + c).$$

Taking expectation completes the proof. □

Lemma 6. *Fix $h > 0$, for arbitrary $u > 0$, define the grid points $t_k := kh$, $k \in \{1, \dots, \lceil \frac{u}{h} \rceil\}$. For arbitrary $\tau > 0$, the following holds*

$$\mathbb{P}\left(\exists k \in \{1, \dots, \lceil \frac{u}{h} \rceil\} : |\widehat{\varphi}(t_k) - \varphi(t_k)| > \tau \left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right) \leq 2\lceil \frac{u}{h} \rceil n^{-\tau^2}.$$

Proof. Lemma 6 is a consequence of the Hoeffding inequality,

$$\begin{aligned} & \mathbb{P}\left(\exists k \leq \lceil \frac{u}{h} \rceil : |\widehat{\varphi}(t_k) - \varphi(t_k)| \geq \tau \left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right) \\ & \leq \sum_{k=1}^{\lceil u/h \rceil} \mathbb{P}\left(|\widehat{\varphi}(t_k) - \varphi(t_k)| \geq \tau \left(\frac{\log n}{n}\right)^{\frac{1}{2}}\right) \leq \sum_{k=1}^{\lceil u/h \rceil} 2 \exp(-2\tau^2(\log n)) = 2\lceil u/h \rceil n^{-\tau^2}. \end{aligned}$$

□

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

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

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

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

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

$$\begin{aligned} |\widehat{\varphi}(v) - \varphi(v)| &\leq |\widehat{\varphi}(v) - \widehat{\varphi}(t_k)| + |\widehat{\varphi}(t_k) - \varphi(t_k)| + |\varphi(t_k) - \varphi(v)| \\ &\leq h(2E[|Y|] + c) + \tau \left(\frac{\log n}{n} \right)^{1/2}. \end{aligned}$$

This is the desired result. □

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 monotonously increasing with respect to τ and $\mathbb{E}[|Y|]$ and monotonously 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 $\kappa > \sqrt{\frac{1}{2} + \frac{1}{K}}$. There exists some $C > 0$ which is monotonously increasing with respect to $\mathbb{E}[Y^2]$ such that for any $u > 0$,*

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

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, $\forall n \geq n_0$,

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

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

6.2 Proof of Theorem 1

We have the decomposition

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

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

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

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

$$\begin{aligned} |\widehat{\varphi}_X(u) - \varphi_X(u)|^2 &= \left| |\widehat{\varphi}(u)|^{\frac{1}{K}} e^{\frac{1}{K}i\operatorname{Im}\widehat{\psi}(u)} - |\varphi(u)|^{\frac{1}{K}} e^{\frac{1}{K}i\operatorname{Im}\psi(u)} \right|^2 \\ &\leq 2 \left| |\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}} \right|^2 + 2 |\varphi(u)|^{\frac{2}{K}} \left| e^{\frac{1}{K}i\operatorname{Im}\widehat{\psi}(u)} - e^{\frac{1}{K}i\operatorname{Im}\psi(u)} \right|^2. \end{aligned} \quad (6.5)$$

First, we consider the case where m lies in $[0, u_{\varepsilon, \kappa}]$, where $u_{\varepsilon, \kappa}$ is defined in (2.2). Consider the first term of (6.2). Using that $x \rightarrow x^{1/K}$ is Hölder continuous for $x > 0$ and the triangle inequality, we derive

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

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

$$\mathbb{E} \left[\int_{-m}^m \left| |\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}} \right|^2 du \right] \leq \int_{-m}^m (\mathbb{E}[|\widehat{\varphi}(u) - \varphi(u)|^2])^{1/K} du \leq C m n^{-\frac{1}{K}}$$

for some positive real valued C .

Next, consider the second term of (6.2). Using the fact that for arbitrary $x, y \in \mathbb{R}$, $|e^{ix} - e^{iy}| \leq |x - y|$ holds, we derive that

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

We observe that, for arbitrary $v \in \mathbb{R}$,

$$\begin{aligned} \frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)} &= \frac{\varphi(v)\widehat{\varphi}'(v) - \widehat{\varphi}(v)\varphi'(v)}{\varphi(v)\widehat{\varphi}(v)} = \frac{\varphi(v)(\widehat{\varphi}'(v) - \varphi'(v)) - \varphi'(v)(\widehat{\varphi}(v) - \varphi(v))}{\varphi(v)\widehat{\varphi}(v)} \\ &= \frac{(\widehat{\varphi}'(v) - \varphi'(v))/\varphi(v) - (\varphi'(v)/\varphi(v)^2)(\widehat{\varphi}(v) - \varphi(v))}{\widehat{\varphi}(v)/\varphi(v)} \\ &= \frac{\left(-\frac{(\widehat{\varphi}(v) - \varphi(v))'}{\varphi(v)} \right)'}{\left(1 - \frac{(\widehat{\varphi}(v) - \varphi(v))}{\varphi(v)} \right)}. \end{aligned} \quad (6.6)$$

By Corollary 1 and by the definition of the event $B_\kappa(m)$ and of $u_{\varepsilon,\kappa}$, we find that on $B_\kappa(m)$,

$$\forall u \in [-m, m] : |\widehat{\varphi}(u) - \varphi(u)| \leq \kappa \sqrt{\log nn}^{-\frac{1}{2}} \leq 1/(1 + \varepsilon)|\varphi(u)|. \quad (6.7)$$

Then, on the event $B_\kappa(m)$, a Neumann series expansion, along with formula (6.2) and formula (6.2) gives for $v \in [-m, m]$,

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

The following representation holds

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

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

$$\begin{aligned} \mathbb{1}_{B_\kappa(m)} \left| \int_0^u \left(\frac{\widehat{\varphi}'(v)}{\widehat{\varphi}(v)} - \frac{\varphi'(v)}{\varphi(v)} \right) dv \right| &\leq \sum_{\ell=0}^{\infty} \frac{1}{\ell + 1} \frac{|\widehat{\varphi}(u) - \varphi(u)|^{\ell+1}}{|\varphi(u)|^{\ell+1}} \\ &\leq \frac{|\widehat{\varphi}(u) - \varphi(u)|}{|\varphi(u)|} \sum_{\ell=0}^{\infty} \frac{(1 + \varepsilon)^{-\ell}}{\ell + 1} = (1 + \varepsilon) \ln(1 + 1/\varepsilon) \frac{|\widehat{\varphi}(u) - \varphi(u)|}{|\varphi(u)|}. \end{aligned}$$

Gathering all the terms together, we have shown that for $m \in [0, u_{\varepsilon,\kappa}]$,

$$\begin{aligned} \mathbb{E} \left[\mathbb{1}_{B_\kappa(m)} \int_{-m}^m |\varphi_X(u)|^2 |e^{\frac{i}{K} \text{Im} \widehat{\psi}(u)} - e^{\frac{i}{K} \psi(u)}|^2 du \right] \\ \leq \frac{c(\varepsilon)}{K^2} \int_{-m}^m |\varphi_X(u)|^2 \frac{\mathbb{E}[|\widehat{\varphi}(u) - \varphi(u)|^2]}{|\varphi(u)|^2} du \leq \frac{c(\varepsilon)}{K^2} \int_{-m}^m \frac{n^{-1}}{|\varphi_X(u)|^{2(K-1)}} du, \end{aligned}$$

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

$$\begin{aligned} \mathbb{E} \left[\mathbb{1}_{B_\kappa(m)^c} \int_{-m}^m |\varphi_X(u)|^2 |e^{\frac{i}{K} \text{Im} \widehat{\psi}(u)} - e^{\frac{i}{K} \psi(u)}|^2 du \right] &\leq 4 \mathbb{P}(B_\kappa(m)^c) \int_{-m}^m |\varphi_X(u)|^2 du \\ &\leq 4 \|\varphi_X\|^2 m n^{-\frac{1}{K}}. \end{aligned}$$

Secondly, we consider the case $m \geq u_{\varepsilon,\kappa}$. The series of inequalities

$$\begin{aligned} |\widehat{\varphi}_X(u) - \varphi_X(u)| &\leq 2|\varphi_X(u)| + \left| |\widehat{\varphi}_X(u)| - |\varphi_X(u)| \right| \\ &= 2|\varphi_X(u)| + \left| |\widehat{\varphi}(u)|^{\frac{1}{K}} - |\varphi(u)|^{\frac{1}{K}} \right| \leq 2|\varphi_X(u)| + |\widehat{\varphi}(u) - \varphi(u)|^{\frac{1}{K}}, \end{aligned}$$

combined with the Jensen inequality, implies that

$$\begin{aligned} \int_{|u| \in [u_{\varepsilon,\kappa}, m]} \mathbb{E} \left[|\widehat{\varphi}_X(u) - \varphi_X(u)|^2 \right] du &\leq 4 \int_{|u| \in [u_{\varepsilon,\kappa}, m]} |\varphi_X(u)|^2 du + 4 \int_{|u| \in [u_{\varepsilon,\kappa}, m]} \mathbb{E} \left[|\widehat{\varphi}(u) - \varphi(u)|^2 \right]^{\frac{1}{K}} du \quad (6.8) \\ &\leq 4 \int_{|u| \in [u_{\varepsilon,\kappa}, m]} |\varphi_X(u)|^2 du + 8n^{-\frac{1}{K}} m. \end{aligned}$$

This completes the proof of of Theorem 1. \square

6.3 Proof of Proposition 1.

Proposition 1 is a direct consequence of Theorem 1. The control of the bias term $\|f - f_m\|$ relies on the following Lemma, that can be found for instance in Butucea and Tsybakov (2008a,b).

Lemma 9. *Consider c, ρ nonnegative real numbers, and β a real such that $2\beta > 1$ if $c = 0$ or $\rho = 0$. Then, for all $m > 0$,*

$$\int_m^\infty (x^2 + 1)^\beta \exp(c|x|^\rho) dx \asymp m^{2\beta+1-\rho} e^{-cm^\rho}$$

$$\int_m^\infty (x^2 + 1)^{-\beta} \exp(-c|x|^\rho) dx \asymp m^{-2\beta+1-\rho} e^{-cm^\rho}.$$

The control of m^* uses the following bounds on $u_{\varepsilon, \kappa}$ and u_a .

- (i) Fix constants β, C, C', C_X and C_f , where $\beta > \frac{1}{2}$ and such that (2.3) is satisfied, and a density $f \in \mathcal{F}(\beta, C, C', C_X, C_f)$. Then, it immediately follows that

$$u_{\varepsilon, \kappa} \geq C_1 \left(\frac{n}{\log(n)} \right)^{\frac{1}{2\beta K}} \quad \text{and} \quad u_a \leq C_2 n^{\frac{1}{2\beta K}}$$

for some positive constants C_1 depending on ε, K and δ and C_2 depending on a .

- (ii) Fix constants $\beta, \rho, c, C, C', C_X$ and C_f , where $\rho > 0$ and such that (2.3) is satisfied, and a density $f \in \mathcal{F}(\beta, \rho, c, C, C', C_X, C_f)$. Then, it immediately follows that

$$u_{\varepsilon, \kappa} \geq C_1 \left(\frac{1}{K} \log \left(\frac{n}{\log(n)} \right) \right)^{\frac{1}{\rho}} \quad \text{and} \quad u_a \leq C_2 \left(\frac{\log(n)}{K} \right)^{\frac{1}{\rho}}$$

for some positive constants C_1 continuously depending on β, ε, K and δ and C_2 depending on a .

Injecting all these approximations in Theorem 1 leads to the desired rates. \square

6.4 Adaptive estimation: Proof of Theorem 2.

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

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

Fix constants $\beta, \rho, c, C, C', C_X$ and C_f and a density f in either $\mathcal{F}(\beta, C, C', C_X, C_f)$ or in $\mathcal{F}(\beta, \rho, c, C, C', C_X, C_f)$ such that (2.3) holds. 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

$$\mathbb{P}(\widehat{m} < m_0) \leq \mathbb{P}(\exists u \in [0, m_0] : |\widehat{\varphi}(u) - \varphi(u)| \geq \kappa \sqrt{\log n} n^{-\frac{1}{2}}) \leq C m_0 n^{-\frac{1}{K}}, \quad (6.9)$$

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,

$$\begin{aligned} \mathbb{P}(\widehat{m} > m_1) &\leq \mathbb{P}(|\widehat{\varphi}(m_1)| > (nK)^{-\frac{1}{2}} + \sqrt{\eta \log(n)/Kn} n^{-\frac{1}{2}}) \\ &\leq \mathbb{P}(|\varphi(m_1) - \widehat{\varphi}(m_1)| > \sqrt{\eta \log(n)/Kn} n^{-\frac{1}{2}}) \leq n^{-\frac{\eta}{K}}. \end{aligned} \quad (6.10)$$

Then, the proof of Theorem 1 implies that

$$\begin{aligned} &\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{m \in [m_0, m_1]}] \\ &\leq \|f - f_{m_0}\|^2 + \frac{C_1}{nK^2} \int_{-m_1}^{m_1} \frac{1}{|\varphi_X(u)|^{2(K-1)}} du + C_2 m_1 n^{-\frac{1}{K}} + \frac{2}{\pi} \int_{|u| \in [u_{\varepsilon, \kappa}, m_1]} |\varphi_X(u)|^2 du, \end{aligned}$$

with a constant C_1 depending on the definition of ε and with C_2 depending on $\|f\| < \overline{C}_f$ and on $\mathbb{E}[X^2] < \overline{C}_X$. Using the fact that $f \in \mathcal{F}(\beta, C, C', C_X, C_f)$ or in $\mathcal{F}(\beta, \rho, c, C, C', C_X, C_f)$ to derive m_0 and m_1 and plugging in m_0 and m_1 gives, for some constant C_1 depending on \overline{C}_f and \overline{C}_X , as well as the choice of ε and δ ,

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{m \in [m_0, m_1]}] \leq C_1 \left(\frac{n}{\log n}\right)^{-\frac{2\beta-1}{2\beta K}} \text{ or } \mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{m \in [m_0, m_1]}] \leq C_1 (\log n)^{\frac{1}{\rho}} n^{-\frac{1}{K}} \quad (6.11)$$

respectively.

Consider now the exceptional set $m \notin [m_0, m_1]$. Applying the triangle inequality and arguing along the same lines as in the proof of Theorem 1, we derive that

$$\begin{aligned} \mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{\widehat{m} < m_0}] &\leq \mathbb{P}(\widehat{m} < m_0) \|f\|^2 + \int_{-m_0}^{m_0} \mathbb{E}[|\widehat{f}_X(u) - f_X(u)|^2] du \\ &\leq \mathbb{P}(\widehat{m} < m_0) \|f\|^2 + \frac{C_1}{nK^2} \int_{-m_0}^{m_0} \frac{1}{|\varphi_X(u)|^2} du + C_2 n^{-\frac{1}{K}} m_0 + \frac{2}{\pi} \int_{|u| \in [u_{\varepsilon, \kappa}, m_0]} |\varphi_X(u)|^2 du. \end{aligned} \quad (6.12)$$

Formula (6.2) and the fact that $\widehat{m} \leq n^{\frac{1}{K}}$ lead to

$$\begin{aligned} \mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{\widehat{m} > m_1}] &\leq \|f - f_{m_1}\|^2 \mathbb{P}(\widehat{m} > m_1) \\ &\quad + 4\mathbb{P}(\widehat{m} > m_1) \int_{-n^{1/K}}^{n^{1/K}} |\varphi_X(u)|^2 du + 8\mathbb{P}(\widehat{m} > m_1)^{\frac{1}{\eta}} \int_{-n^{1/K}}^{n^{1/K}} n^{-\frac{1}{K}} du. \end{aligned} \quad (6.13)$$

Combining (6.4) and (6.4) together with (6.4) and (6.4) and plugging in m_0 and m_1 , we have shown that for a positive constant C_2 depending on \overline{C}_f , \overline{C}_X and the choice of ε and δ and η ,

$$\mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{m \notin [m_0, m_1]}] \leq C_2 \left(\frac{n}{\log n}\right)^{-\frac{2\beta-1}{2\beta K}} \text{ or } \mathbb{E}[\|f - \widehat{f}_{\widehat{m}}\|^2 \mathbf{1}_{m \notin [m_0, m_1]}] \leq C_2 (\log n)^{\frac{1}{\rho}} n^{-\frac{1}{K}} \quad (6.14)$$

respectively. Putting (6.4) and (6.4) together, we have shown that for a constant \mathcal{C} depending on the choice of η and ε , on \bar{C}_f and \bar{C}_X ,

$$\mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2](n/\log n)^{\frac{2\beta-1}{2\beta K}} \leq \mathcal{C} \quad \text{or} \quad \mathbb{E}[\|f - \hat{f}_{\hat{m}}\|^2](\log n)^{-\frac{1}{\rho}} n^{\frac{1}{K}} \leq \mathcal{C}.$$

Taking the supremum in $\beta, \rho, c, C, C', C_X$ and C_f on \bar{I} gives the statement of the theorem. \square

References

- [1] Bøgsted, M. and Pitts, S. M. (2010). Decomposing random sums: a nonparametric approach. *Annals of the Institute of Statistical Mathematics*, 62(5), 855-872.
- [2] Buchmann, B. and Grübel, R. (2003). Decomposing: an estimation problem for Poisson random sums. *The Annals of Statistics*, **31**, 1054–1074.
- [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] Comte, F. and Genon-Catalot, V. (2009). Nonparametric estimation for pure jump Lévy processes based on high frequency data. *Stochastic Processes and their Applications*, **119**, 4088–4123
- [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] 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.
- [10] Duval, C. (2013a) Nonparametric estimation of a renewal reward process from discrete data. *Mathematical Methods of Statistics*, **22**, 28-56.
- [11] Duval, C. (2013b) Density estimation for compound Poisson processes from discrete data. *Stochastic Processes and their Applications*, **123**, 3963–3986.
- [12] Duval, C. (2014) When is it no longer possible to estimate a compound Poisson process? *Electronic Journal of Statistics* **8**, 274–301.
- [13] van Es, B., Gugushvili, S. and Spreij, P. (2007). A kernel type nonparametric density estimator for decomposing, *Bernoulli*, **13**, 672–694.
- [14] Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Processes and their Applications* **124**, 730–758.
- [15] Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution, *Electronic Journal of Statistics*, **8**, 2879-2904.

- [16] Meister, A. (2007). Optimal convergence rates for density estimation from grouped data. *Statistics & probability letters*, **77**(11), 1091–1097.
- [17] Neumann, M. and Reiß, M. (2009) Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli* **15**, 223–248.