Sharp optimality and some effects of dominating bias in density deconvolution

Cristina BUTUCEA[1,2] and Alexandre B. TSYBAKOV[1]
[1] Université Paris VI, [2] Université Paris X

March 30, 2004

Abstract

We consider estimation of the common probability density f of i.i.d. random variables X_i that are observed with an additive i.i.d. noise. We assume that the unknown density f belongs to a class $\mathcal{A}$ of densities whose characteristic function is described by the exponent $\exp(-\alpha|u|^r)$ as $|u| \to \infty$, where $\alpha > 0$, $r > 0$. The noise density is supposed to be known and such that its characteristic function decays as $\exp(-\beta|u|^s)$, as $|u| \to \infty$, where $\beta > 0$, $s > 0$. Assuming that $r < s$, we suggest a kernel type estimator that is optimal in sharp asymptotical minimax sense on $\mathcal{A}$ simultaneously under the pointwise and the $\mathbb{L}_2$-risks. The variance of this estimator turns out to be asymptotically negligible w.r.t. its squared bias. For $r < s/2$ we construct a sharp adaptive estimator of f. We discuss some effects of dominating bias, such as superefficiency of minimax estimators.

Mathematics Subject Classifications: 62G05, 62G20
Key Words: Deconvolution, nonparametric density estimation, infinitely differentiable functions, exact constants in nonparametric smoothing, minimax risk, adaptive curve estimation.
Short title: Sharp optimality in density deconvolution

1 Introduction

Assume that one observes $Y_1, \dots, Y_n$ in the model

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \dots, n,$$

where X_i are i.i.d. random variables with an unknown probability density f w.r.t. the Lebesgue mesure on $\mathbb{R}$, the random variables ε_i are i.i.d. with known probability density f^ε w.r.t. the Lebesgue mesure on $\mathbb{R}$, and $(\varepsilon_1, \dots, \varepsilon_n)$ is independent of $(X_1, \dots, X_n)$. The deconvolution problem that we consider here is to estimate f from observations $Y_1, \dots, Y_n$.

Denote by $f^Y = f * f^\varepsilon$ the density of the variables Y_i, where $*$ is the convolution sign. Let Φ^Y, Φ^X and Φ^ε be the characteristic functions of random variables Y_i, X_i and ε_i, respectively. For an integrable function $g : \mathbb{R} \to \mathbb{R}$, define the Fourier transform

$$\Phi^g(u) = \int g(x) \exp(ixu) du.$$

We assume that the unknown density f belongs to the class of functions

$$\mathcal{A}_{\alpha,r}(L) = \{f \text{ is a probability density on } \mathbb{R} \text{ and } \int \left|\Phi^f(u)\right|^2 \exp(2\alpha|u|^r) du \le 2\pi L\},$$

where $\alpha > 0$, $r > 0$, $L > 0$ are finite constants. The classes of densities of this type have been studied by many authors starting from Ibragimov and Hasminskii (1983). For a recent overview see Belitser and Levit (2001) and Artiles (2001).

We suppose also in most of the results that the characteristic function of noise ε_i satisfies the following assumption.

Assumption (N). *There exist constants $u_0 > 0$, $\beta > 0$, $s > 0$, $b_{\min} > 0$, $b_{\max} > 0$ and $\gamma, \gamma' \in \mathbb{R}$ such that*

$$b_{\min}|u|^\gamma \exp(-\beta\,|u|^s) \le |\Phi^\varepsilon(u)| \le b_{\max}|u|^{\gamma'} \exp(-\beta\,|u|^s) \tag{1}$$

for $|u| \ge u_0$.

Many important probability densities belong to the class $\mathcal{A}_{\alpha,r}(L)$ with some α, r, L or have the characteristic function satisfying (1). All such densities are infinitely many times differentiable on $\mathbb{R}$. Examples include normal, Cauchy and general stable laws, Student, logistic, extreme value distributions and other, as well as their mixtures and convolutions. Note that in these examples the values r and/or s are less or equal to 2. Although the densities with $r > 2, s > 2$ are in principle conceivable, they are difficult to express in a closed form, and the set of such densities does not contain statistically famous representatives. This remark concerns especially the noise density f^ε that should be explicitly known. Therefore, without a meaningful loss, we will sometimes restrict our study to the case $0 < s \le 2$.

For any estimator $\widehat{f}_n$ of f define the maximal pointwise risk over the class $\mathcal{A}_{\alpha,r}(L)$ for any fixed $x \in \mathbb{R}$ by

$$R_n(x, \widehat{f}_n, \mathcal{A}_{\alpha,r}(L)) = \sup_{f \in \mathcal{A}_{\alpha,r}(L)} E_f \left[\left|\widehat{f}_n(x) - f(x)\right|^2\right]$$

and the maximal $\mathbb{L}_2$-risk

$$R_n(\mathbb{L}_2, \widehat{f}_n, \mathcal{A}_{\alpha,r}(L)) = \sup_{f \in \mathcal{A}_{\alpha,r}(L)} E_f \left[\|\widehat{f}_n - f\|_2^2\right],$$

where $E_f(\cdot)$ is the expectation with respect to the joint distribution P_f of $Y_1, \ldots, Y_n$, when the underlying probability density of X_i's is f, and $\|\cdot\|_2$ stands for the $\mathbb{L}_2(\mathbb{R})$-norm. (In what follows we use the notation $\mathbb{L}_p(\mathbb{R})$, in general, for the $\mathbb{L}_p$-spaces of complex valued functions on $\mathbb{R}$.)

The asymptotics of optimal estimators differ significantly for the cases $r < s$, $r = s$ and $r > s$. If $r < s$ the variance of the optimal estimator is asymptotically negligible w.r.t. the bias, while for $r > s$ the bias is asymptotically negligible w.r.t. the variance. In this paper we consider the *bias dominated case*, i.e. we assume that $r < s$. The setting with dominating variance will be treated in another paper.

The problems of density deconvolution with dominating bias were historically the first ones studied in the literature [cf. Ritov (1987), Stefanski and Carroll (1990), Carroll and Hall (1988), Zhang (1990), Fan (1991a,b), Masry (1991), Efromovich (1997)], motivated by the importance of deconvolution with gaussian noise. These papers consider, in particular, the noise distributions satisfying (1), but the densities f belonging to finite smoothness classes, such as Hölder or Sobolev ones, where the estimation of f is harder than for the class $\mathcal{A}_{\alpha,r}(L)$. In this framework they show that optimal rates of convergence are as a power of $\log n$ which suggests that essentially there is no hope to recover f with a reasonably small error for reasonable sample sizes. This conclusion is often interpreted as a general pessimistic message about the gaussian deconvolution problem. Note, however, that such minimax results are obtained for the least favorable densities in Hölder or Sobolev classes. Often the underlying density is much nicer (for instance, it belongs to $\mathcal{A}_{\alpha,r}(L)$, as the popular densities mentioned above), and the estimation can be significantly improved, as we show below: the optimal rates of convergence are in fact faster than any power of $\log n$.

Pensky and Vidakovic (1999) studied density deconvolution in the classes of densities that are somewhat smaller than $\mathcal{A}_{\alpha,r}(L)$ (including an additional restriction on the tails of f) and with the noise satisfying (1). They analyzed the rates of convergence of wavelet deconvolution estimators, restricting their attention to the $\mathbb{L}_2$-risk. Our results imply that the rates achieved by their estimators are not optimal on $\mathcal{A}_{\alpha,r}(L)$ and that the optimal rates can be attained by a simpler and more traditional kernel deconvolution method with suitably chosen parameters. We will show that our method attains not only the optimal rates but also the best asymptotic constants (i.e. is sharp optimal). Moreover, we will prove that the proposed estimator is sharp optimal simultaneously under the $\mathbb{L}_2$-risk and under the pointwise risk and that it is sharp adaptive to the parameters α, r, L in some cases.

The most difficult part of our results is the construction of minimax lower bounds. The technique that we develop might be useful to get lower bounds for similar "2 exponents" type settings in other inverse problems. To our knowledge, except for the case $r = s = 1$ treated by Golubev and Khasminskii (2001), Tsybakov (2000) and Cavalier, Golubev, Lepski and Tsybakov (2003), such lower bounds are not available even for the Gaussian white noise (or sequence space) deconvolution model, although some upper bounds are known (cf. Ermakov (1989), Efromovich and Koltchinskii (2001)).

Finally, we mention publications on adaptive deconvolution under Assumption (N) or its analogs. They deal with the problems that are somewhat different from ours. Efromovich (1997) considered the problem of deconvolution where the densities f and f^ε are both periodic on $[0, 2\pi]$, f^ε satisfies an analog of Assumption (N) expressed in terms of Fourier coefficients and f belongs to a class of periodic functions of Sobolev type. He proposed sharp adaptive estimators with logarithmic rates which are optimal for that framework, as discussed above. Adaptive deconvolution in a gaussian white noise model had been studied by Goldenshluger (1998). He worked under the Assumption (N) on the Fourier transform of the convolution kernel or under the assumption that it decreases as a power of u, as $|u| \to \infty$, but he assumed that the function f to estimate belongs to a Sobolev class with unknown parameters. He proposed a rate adaptive estimator under the pointwise risk.

2 The estimator, its bias and variance

Consider the following kernel estimator of f:

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} K_n\left(\frac{Y_i - x}{h_n}\right), \tag{2}$$

where $h_n > 0$ is a bandwidth and K_n is the function on $\mathbb{R}$ defined as the inverse Fourier transform of

$$\Phi^{K_n}(u) = \frac{I(|u| \leq 1)}{\Phi^\varepsilon(u/h_n)}. \tag{3}$$

Here and later $I(\cdot)$ denotes the indicator function. The function K_n is called kernel, but unlike the usual Parzen-Rosenblatt kernels, it depends on n.

For the existence of K_n it is enough that $\Phi^{K_n} \in \mathbb{L}_2(\mathbb{R})$ (and thus $\Phi^{K_n} \in \mathbb{L}_1(\mathbb{R})$). This holds under mild assumptions. For example, in view of the continuity property of characterictic functions, the assumption that $\Phi^\varepsilon(u) \neq 0$ for all $u \in \mathbb{R}$ is sufficient to have $\Phi^{K_n} \in \mathbb{L}_2(\mathbb{R})$. Moreover, the condition $\Phi^{K_n} \in \mathbb{L}_2(\mathbb{R})$ implies that the kernel K_n is real-valued. In fact, under this condition we have $\Phi^{K_n}(u) = \Phi^\varepsilon(-u/h_n)V_n(u)$ for almost all $u \in \mathbb{R}$, where $V_n(u) = I(|u| \leq 1)/|\Phi^\varepsilon(u/h_n)|^2$ is an even real-valued function belonging to $\mathbb{L}_1(\mathbb{R})$ and $\Phi^\varepsilon(-u/h_n)$ (the complex conjugate of $\Phi^\varepsilon(u/h_n)$) is the Fourier transform of real-valued function $t \mapsto h_n f^\varepsilon(-h_n t)$. This implies that K_n is a convolution of two real-valued functions.

The estimator (2) belongs to the family of kernel deconvolution estimators studied in many papers starting from Stefanski and Carroll (1990), Carroll and Hall (1988) and Zhang (1990). It can be also deduced from a unified approach to construction of estimators in statistical inverse problems (Ruymgaart (1993)).

The following proposition establishes upper bounds on the pointwise and the $\mathbb{L}_2$ *bias terms*, i.e. on the quantities $|E_f \hat{f}_n(x) - f(x)|^2$ and $\|E_f \hat{f}_n - f\|_2^2$.

Proposition 1 *Let $f \in \mathcal{A}_{\alpha,r}(L)$, $\alpha > 0, r > 0, L > 0$ and assume that $\Phi^{K_n} \in \mathbb{L}_2(\mathbb{R})$ for any $h_n > 0$. Then the squared bias of $\hat{f}_n(x)$ is bounded as follows*

$$\sup_{x \in \mathbb{R}} \left| E_f \hat{f}_n(x) - f(x) \right|^2 \leq \frac{L}{2\pi\alpha r} h_n^{r-1} \exp\left(-\frac{2\alpha}{h_n^r}\right)(1 + o(1)),$$

as $h_n \to 0$, while the bias term of the $\mathbb{L}_2$-risk satisfies

$$\|E_f\hat{f}_n - f\|_2^2 \leq L\exp\left(-\frac{2\alpha}{h_n^r}\right)$$

for every $h_n > 0$.

Proof. For the pointwise bias we have

$$\begin{aligned}
\left|E_f\hat{f}_n(x) - f(x)\right|^2 &= \left|\left(\frac{1}{h_n}K_n\left(\frac{\cdot}{h_n}\right) * f^Y(\cdot)\right)(x) - f(x)\right|^2 \\
&= \left|\frac{1}{2\pi}\int\left[\Phi^{K_n}(uh_n)\Phi^Y(u) - \Phi^X(u)\right]\exp(-iux)du\right|^2 \\
&\leq \frac{1}{(2\pi)^2}\left(\int I(|uh_n| > 1)|\Phi^X(u)|du\right)^2.
\end{aligned}$$

Applying the Cauchy-Schwarz inequality and the assumption that f belongs to $\mathcal{A}_{\alpha,r}(L)$ we get

$$\begin{aligned}
&\left|E_f\hat{f}_n(x) - f(x)\right|^2 \\
\leq\ & \frac{1}{(2\pi)^2}\int_{|u|>1/h_n}\exp(-2\alpha|u|^r)du\int_{|u|>1/h_n}|\Phi^X(u)|^2\exp(2\alpha|u|^r)du \qquad (4)\\
\leq\ & \frac{L}{2\pi}\int_{|u|>1/h_n}\exp(-2\alpha|u|^r)du
\end{aligned}$$

which together with Lemma 6 yields the first inequality of the Proposition. To prove the second inequality, we apply the Plancherel formula and get

$$\begin{aligned}
\|E_f\hat{f}_n - f\|_2^2 &= \left\|\frac{1}{h_n}E_fK_n\left(\frac{Y_1 - \cdot}{h_n}\right) - f(\cdot)\right\|_2^2 \\
&= \frac{1}{2\pi}\int\left|\Phi^{K_n}(uh_n)\Phi^Y(u) - \Phi^X(u)\right|^2du \\
&= \frac{1}{2\pi}\int I(|uh_n| > 1)|\Phi^X(u)|^2du \\
&\leq \frac{\exp(-2\alpha/h_n^r)}{2\pi}\int_{|u|>1/h_n}\left|\Phi^X(u)\right|^2\exp(2\alpha\,|u|^r)du. \qquad (5)
\end{aligned}$$

□

The next proposition gives upper bounds on the pointwise and the $\mathbb{L}_2$ *variance terms* defined as

$$Var_f\hat{f}_n(x) = E_f\left[|\hat{f}_n(x) - E_f\hat{f}_n(x)|^2\right] \quad \text{and} \quad Var_{f,2}\hat{f}_n = E_f\left[\|\hat{f}_n - E_f\hat{f}_n\|_2^2\right]$$

respectively.

Proposition 2 *Let the left inequality in (1) hold and* $\Phi^\varepsilon(u) \neq 0, \forall\ u \in \mathbb{R}$. *Then, for any density* f *such that* $\sup_{x\in\mathbb{R}} f(x) \le f^* < \infty$, *the pointwise variance of the estimator* $\hat{f}_n(x)$ *is bounded as follows*

$$\sup_{x\in\mathbb{R}} Var_f \hat{f}_n(x) = \sup_{x\in\mathbb{R}} E_f\left[|\hat{f}_n(x) - E_f\hat{f}_n(x)|^2\right] \le \frac{f_* h_n^{s-2\gamma-1}}{2\pi\beta s b_{\min}^2 n} \exp\left(\frac{2\beta}{h_n^s}\right)(1+o(1)), \tag{6}$$

as $h_n \to 0$, *and, for an arbitrary density* f, *the variance term of the* $\mathbb{L}_2$*-risk satisfies*

$$Var_{f,2}\hat{f}_n = E_f\left[\|\hat{f}_n - E_f\hat{f}_n\|_2^2\right] \le \frac{h_n^{s-2\gamma-1}}{2\pi\beta s b_{\min}^2 n} \exp\left(\frac{2\beta}{h_n^s}\right)(1+o(1)) \tag{7}$$

as $h_n \to 0$.

Proof. Define

$$K_{2,n}(x) = \frac{1}{h_n} K_n^2\left(\frac{x}{h_n}\right).$$

For the pointwise variance we write

$$\begin{aligned} Var_f \hat{f}_n(x) &= \frac{1}{n} E_f\left[\left|\frac{1}{h_n}K_n\left(\frac{Y_1 - x}{h_n}\right) - E_f\left[\frac{1}{h_n}K_n\left(\frac{Y_1-x}{h_n}\right)\right]\right|^2\right] \\ &\le \frac{1}{nh_n}(K_{2,n} * f^Y)(x) \\ &\le \frac{f_*}{nh_n}\|K_n\|_2^2, \end{aligned} \tag{8}$$

where we used the fact that the convolution density $f^Y = f * f^\varepsilon$ is uniformly bounded by f_*. Applying the Plancherel formula and using (1) and (64) of Lemma 6 in the Appendix we get

$$\begin{aligned} \|K_n\|_2^2 &= \frac{h_n}{2\pi}\int_{|u|\le 1/h_n} |\Phi^\varepsilon(u)|^{-2} du \\ &\le \frac{h_n}{2\pi b_{\min}^2}\int_{u_0 \le |u| \le 1/h_n} |u|^{2\gamma}\exp(2\beta|u|^s)du + \frac{h_n}{2\pi}\int_{|u|\le u_0}|\Phi^\varepsilon(u)|^{-2}du \\ &\le \frac{h_n}{\pi b_{\min}^2}\int_0^{1/h_n} u^{2\gamma}\exp(2\beta u^s)du + O(h_n) \\ &= \frac{h_n^{s-2\gamma}}{2\pi b_{\min}^2 \beta s}\exp\left(\frac{2\beta}{h_n^s}\right)(1+o(1)), \quad h_n \to 0. \end{aligned} \tag{9}$$

This and (8) imply (6). In a similar way

$$\begin{aligned} Var_{f,2}\hat{f}_n &\le \frac{1}{nh_n}\int (K_{2,n} * f^Y)(x)dx \\ &= \frac{1}{nh_n}\|K_n\|_2^2, \end{aligned}$$

and in view of (9) we obtain (7). $\square$

Clearly, the bounds of Proposition 2 can be applied to $f \in \mathcal{A}_{\alpha,r}(L)$ with, for example,

$$f_* = \sup_{f\in\mathcal{A}_{\alpha,r}(L)} \sup_{x\in\mathbb{R}} |f(x)|.$$

This value is finite and can be taken as in Lemma 5 of the Appendix.

3 Optimal bandwidths and upper bounds for the risks

Propositions 1 and 2 lead to upper bounds for pointwise and $\mathbb{L}_2$ risks that can be minimized in h_n. In this section we give an asymptotic approximation for the result of such a minimization assuming that $r < s$. The corresponding solutions h_n will be called optimal bandwidths. Note that here we consider only optimization within a given class of estimators, moreover we minimize upper bounds on the risks and not the exact risks. However, this turns out to be precise enough in asymptotical sense: in the next section we will show that the estimator $\hat{f}_n$ with optimal bandwidth is sharp minimax over all possible estimators.

Decomposition of the mean squared error of the kernel estimator into bias and variance terms and application of Propositions 1 and 2 yields

$$\begin{aligned} E_f\left[\left|\hat{f}_n(x) - f(x)\right|^2\right] &= \left|E_f\hat{f}_n(x) - f(x)\right|^2 + Var_f\hat{f}_n(x) \\ &\le \frac{L}{2\pi\alpha r} h_n^{r-1}\exp\left(-\frac{2\alpha}{h_n^r}\right) + \frac{f_*}{2\pi\beta s b_{\min}^2}\frac{h_n^{s-2\gamma-1}}{n}\exp\left(\frac{2\beta}{h_n^s}\right). \end{aligned}$$

We now minimize the last expression in h_n. Clearly, the minimizer $h_n = \tilde{h}_n$ tends to 0, as $n \to \infty$. Taking derivatives with respect to h_n and neglecting the smaller terms lead us to the equation for optimal bandwidth

$$\frac{Lb_{\min}^2}{f_*} n\tilde{h}_n^{2\gamma}(1+o(1)) = \exp\left(\frac{2\alpha}{\tilde{h}_n^r} + \frac{2\beta}{\tilde{h}_n^s}\right), \tag{10}$$

(asymptotics are taken as $\tilde{h}_n \to 0$, $n \to \infty$). Taking logarithms in the above equation we obtain that the optimal bandwidth $\tilde{h}_n$ is a solution in h of the equation

$$-2\gamma\log h + \frac{2\alpha}{h^r} + \frac{2\beta}{h^s} = \log n + C(1+o(1)), \tag{11}$$

Here and in what follows we denote by C constants with values in $\mathbb{R}$ that can be different on different occasions. For the bandwidth $h = \tilde{h}_n$ satisfying (10) and (11) we can write

$$\begin{aligned} \tilde{h}_n^{r-1}\exp\left(-\frac{2\alpha}{\tilde{h}_n^r}\right) &= C(1+o(1))\frac{\tilde{h}_n^{r-2\gamma-1}}{n}\exp\left(\frac{2\beta}{\tilde{h}_n^s}\right) \\ &= C(1+o(1))\tilde{h}_n^{r-s}\frac{\tilde{h}_n^{s-2\gamma-1}}{n}\exp\left(\frac{2\beta}{\tilde{h}_n^s}\right), \end{aligned}$$

with some constant $C > 0$. This proves that, for the optimal bandwidth, the bias term dominates the variance term whenever $r < s$. (Strictly speaking, here we consider upper bounds on the bias and variance terms and not precisely these terms.)

Similarly, for the $\mathbb{L}_2$-risk we get

$$\begin{aligned} E_f\left[\|\hat{f}_n - f\|_2^2\right] &= \|E_f\hat{f}_n - f\|_2^2 + Var_{f,2}\hat{f}_n \\ &\leq L\exp\left(-\frac{2\alpha}{h_n^r}\right) + \frac{1}{2\pi\beta s b_{\min}^2}\frac{h_n^{s-2\gamma-1}}{n}\exp\left(\frac{2\beta}{h_n^s}\right), \end{aligned}$$

and the minimizer $h_n = h_n(\mathbb{L}_2)$ of the last expression is a solution in h of the equation

$$(r - 2\gamma - 1)\log h + \frac{2\alpha}{h^r} + \frac{2\beta}{h^s} = \log n + C(1 + o(1)). \tag{12}$$

Now, this equation implies

$$\begin{aligned} \exp\left(-\frac{2\alpha}{h_n^r(\mathbb{L}_2)}\right) &= C(1+o(1))\frac{h_n^{r-2\gamma-1}(\mathbb{L}_2)}{n}\exp\left(\frac{2\beta}{h_n^s(\mathbb{L}_2)}\right) \\ &= C(1+o(1))h_n^{r-s}(\mathbb{L}_2)\frac{h_n^{s-2\gamma-1}(\mathbb{L}_2)}{n}\exp\left(\frac{2\beta}{h_n^s(\mathbb{L}_2)}\right), \end{aligned}$$

for some constant $C > 0$. This proves that also for the $\mathbb{L}_2$-risk the bias term dominates the variance term whenever $r < s$.

Thus we obtain two different equations (11) and (12) that define optimal bandwidths for pointwise and $\mathbb{L}_2$ risks respectively, and in both cases the bias terms are asymptotically dominating.

In fact, we can obtain the same results using a single bandwidth defined as follows. Denote by $h_* = h_*(n)$ the unique solution of the equation

$$\frac{2\beta}{h_*^s} + \frac{2\alpha}{h_*^r} = \log n - (\log\log n)^2, \tag{13}$$

(in what follows we will assume w.l.o.g. that $n \geq 3$ to ensure that $\log n > (\log\log n)^2$). Lemma 8 in the Appendix implies that, both for the pointwise and the $\mathbb{L}_2$ loss, the bias terms of the estimator $\hat{f}_n$ with bandwidth h_* given by (13) are of the same order as those corresponding to bandwidths $\tilde{h}_n$ and $h_n(\mathbb{L}_2)$, while the variance terms corresponding to (13) are asymptotically smaller. Thus, the pointwise risk and the $\mathbb{L}_2$ risk of the estimator $\hat{f}_n$ with bandwidth h_* given by (13) are asymptotically of the same order as those for estimators $\hat{f}_n$ with optimal bandwidths $\tilde{h}_n$ and $h_n(\mathbb{L}_2)$ respectively.

Note that, in fact, h_* is better than both bandwidths $\tilde{h}_n$ and $h_n(\mathbb{L}_2)$ in the variance terms, but these terms are asymptotically negligible w.r.t. the bias ones (cf. Lemma 8). Therefore, the improvement does not appear in the main term of the asymptotics. Note also that the sequence $(\log\log n)^2$ in (13) can be replaced by a sequence satisfying $b_n = o((\log n)^{1-r/s})$, $b_n/\log\log n \to \infty$ and the above argument remains valid (cf. the proof of Lemma 8).

Calculating the upper bounds for bias terms of the estimator $\hat{f}_n$ with bandwidth (13) we get the following asymptotical upper bounds for its pointwise and $\mathbb{L}_2$ risks respectively:

$$\varphi_n^2 = \frac{L}{2\pi\alpha r} h_*^{r-1} \exp\left(-\frac{2\alpha}{h_*^r}\right) = \frac{L}{2\pi\alpha r}\left(\frac{\log n}{2\beta}\right)^{(1-r)/s} \exp\left(-\frac{2\alpha}{h_*^r}\right)(1+o(1)) \quad (14)$$

and

$$\varphi_n^2(\mathbb{L}_2) = L \exp\left(-\frac{2\alpha}{h_*^r}\right). \quad (15)$$

The above remarks can be summarized as follows.

Theorem 1 *Let $\alpha > 0, L > 0, 0 < r < s < \infty$, let the left inequality in (1) hold and $\Phi^\varepsilon(u) \neq 0, \forall\, u \in \mathbb{R}$. Then the kernel estimator $\hat{f}_n$ with bandwidth defined by (13) satisfies the following pointwise and $\mathbb{L}_2$-risk bounds*

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} R_n(x, \hat{f}_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} \leq 1, \quad (16)$$

$$\limsup_{n\to\infty} R_n(\mathbb{L}_2, \hat{f}_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) \leq 1, \quad (17)$$

where the rates φ_n and $\varphi_n(\mathbb{L}_2)$ are given in (14) and (15).

Consider now the case $r \leq s/2$ that is of a particular interest. It covers the situation where the noise density f^ε is gaussian ($s = 2$) and the underlying density f admits the analytic continuation into a strip of the complex plane ($r = 1$), as it is the case for the statistically famous densities mentioned in the introduction. The classes with $r < 1$ are even larger. It is easy to see that

$$\varphi_n^2 = \begin{cases} \frac{L}{2\pi\alpha r}\left(\frac{\log n}{2\beta}\right)^{(1-r)/s} \exp\left(-2\alpha\left(\frac{\log n}{2\beta}\right)^{r/s}\right)(1+o(1)), & \text{if } r < s/2, \\ \frac{L}{2\pi\alpha r}\left(\frac{\log n}{2\beta}\right)^{(1-r)/s} \exp\left(-2\alpha\sqrt{\frac{\log n}{2\beta}} + \frac{\alpha^2}{\beta}\right)(1+o(1)), & \text{if } r = s/2 \end{cases} \quad (18)$$

and

$$\varphi_n^2(\mathbb{L}_2) = \begin{cases} L \exp\left(-2\alpha\left(\frac{\log n}{2\beta}\right)^{r/s}\right)(1+o(1)), & \text{if } r < s/2, \\ L \exp\left(-2\alpha\sqrt{\frac{\log n}{2\beta}} + \frac{\alpha^2}{\beta}\right)(1+o(1)), & \text{if } r = s/2. \end{cases} \quad (19)$$

The bandwidth (13) depends on the parameters α, r of the class $\mathcal{A}_{\alpha,r}(L)$ that are not known in practice. However, it is possible to construct an adaptive estimator that does not depend on these parameters and that attains the same asymptotic behavior as in Theorem 1 both for pointwise and $\mathbb{L}_2$ risks when $r < s/2$. Define the set of parameters

$$\Theta = \{(\alpha, L, r) : \alpha > 0, L > 0, 0 < r < s/2\}.$$

Note that the parameters s and β are supposed to be known since they characterize the known density of noise f^ε.

Theorem 2 *Suppose that the left inequality in (1) holds and* $\Phi^{\varepsilon}(u) \neq 0, \forall\, u \in \mathbb{R}$. *Let* $f_n^{\mathbf{a}}$ *be kernel estimator defined in* (2) *with bandwidth* $h_n = h_n^{\mathbf{a}}$ *defined by*

$$h_n^{\mathbf{a}} = \left(\frac{\log n}{2\beta} - \sqrt{\frac{\log n}{2\beta}} \right)^{-1/s} \tag{20}$$

for n *large enough so that* $\log n/(2\beta) > 1$. *Then, for all* $(\alpha, L, r) \in \Theta$,

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} R_n(x, f_n^{\mathbf{a}}, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} \le 1,$$

and

$$\limsup_{n\to\infty} R_n(\mathbb{L}_2, f_n^{\mathbf{a}}, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) \le 1,$$

where the rates φ_n *and* $\varphi_n(\mathbb{L}_2)$ *are defined in* (14) *and* (15) *(and, more particularly, satisfy* (18) *and* (19) *with* $r < s/2$*).*

Proof. Since $r/s < 1/2$, we have $-\left(\frac{\log n}{2\beta} - \sqrt{\frac{\log n}{2\beta}}\right)^{r/s} > -\frac{\beta}{2\alpha}\sqrt{\frac{\log n}{2\beta}}$ for n large enough, and thus

$$\exp\left(-\frac{2\alpha}{(h_n^{\mathbf{a}})^r}\right) \ge \exp\left(-\beta\sqrt{\frac{\log n}{2\beta}}\right).$$

On the other hand,

$$\frac{1}{n}\exp\left(\frac{2\beta}{(h_n^{\mathbf{a}})^s}\right) = \exp\left(-2\beta\sqrt{\frac{\log n}{2\beta}}\right).$$

Therefore, the ratio of the bias term of $f_n^{\mathbf{a}}$ to the variance term of $f_n^{\mathbf{a}}$ both for the pointwise risk and for the $\mathbb{L}_2$-risk is bounded from below by

$$(\log n)^b \exp\left(\beta\sqrt{\frac{\log n}{2\beta}}\right)$$

for some $b \in \mathbb{R}$. This expression tends to ∞ as $n \to \infty$. Thus, the variance terms are asymptotically negligible w.r.t. the bias terms. It remains to check that the bias terms of $f_n^{\mathbf{a}}$ for both risks are asymptotically bounded by φ_n^2 and $\varphi_n^2(\mathbb{L}_2)$ respectively.

In view of Proposition 1, for n large enough the bias term of $f_n^{\mathbf{a}}$ for the pointwise risk is bounded from above by

$$\begin{aligned}
&\frac{L}{2\pi\alpha r}(h_n^{\mathbf{a}})^{r-1}\exp\left(-2\alpha\left(\frac{\log n}{2\beta}\right)^{r/s}\left[1-\left(\frac{\log n}{2\beta}\right)^{-1/2}\right]^{r/s}\right)\\
&\le \frac{L}{2\pi\alpha r}\left(\frac{\log n}{2\beta}\right)^{(1-r)/s}\exp\left(-2\alpha\left(\frac{\log n}{2\beta}\right)^{r/s} + c\left(\frac{\log n}{2\beta}\right)^{r/s-1/2}\right)(1+o(1))\\
&= \varphi_n^2(1+o(1)),
\end{aligned}$$

where $c > 0$ is a constant and we have used (18) with $r < s/2$ for the last equality. Similarly, for n large enough the bias term of $f_n^{\mathbf{a}}$ for the $\mathbb{L}_2$-risk is bounded from above by

$$\begin{aligned} & L \exp\left(-2\alpha \left(\frac{\log n}{2\beta}\right)^{r/s} \left[1 - \left(\frac{\log n}{2\beta}\right)^{-1/2}\right]^{r/s}\right) \\ & \leq L \exp\left(-2\alpha \left(\frac{\log n}{2\beta}\right)^{r/s} + c\left(\frac{\log n}{2\beta}\right)^{r/s-1/2}\right) = \varphi_n^2(\mathbb{L}_2)(1 + o(1)), \end{aligned}$$

where $c > 0$ and we have used (19) with $r < s/2$ for the last equality. □

If $r = s/2$, adaptation to (α, L) is still possible via a procedure similar to that of Theorem 2, but it does not attain the exact constant, as shows the following result. Introduce the set

$$\Theta_0 = \{(\alpha, L) : 0 < \alpha \leq \alpha_0, L > 0\},$$

where $\alpha_0 > 0$ is a constant.

Theorem 3 *Suppose that the left inequality in (1) holds and $\Phi^{\varepsilon}(u) \neq 0, \forall\ u \in \mathbb{R}$. Let $f_n^{\mathbf{a}}$ be the kernel estimator defined in* (2) *with bandwidth $h_n = h_n^{\mathbf{a}}$ defined by*

$$h_n^{\mathbf{a}} = \left(\frac{\log n}{2\beta} - \frac{A}{\beta}\sqrt{\frac{\log n}{2\beta}}\right)^{-1/s}$$

where $A > \alpha_0$ and n is large enough so that $\log n/(2\beta) > (A/\beta)^2$. Then for $r = s/2$ and for all $(\alpha, L) \in \Theta_0$,

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}} R_n(x, f_n^{\mathbf{a}}, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} \leq \exp\left(\frac{\alpha A}{\beta} - \frac{\alpha^2}{\beta}\right), \tag{21}$$

$$\limsup_{n\to\infty} R_n(\mathbb{L}_2, f_n^{\mathbf{a}}, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) \leq \exp\left(\frac{\alpha A}{\beta} - \frac{\alpha^2}{\beta}\right), \tag{22}$$

where the rates φ_n and $\varphi_n(\mathbb{L}_2)$ are given in (18) *and* (19).

Proof. It is easily checked that the bias exponent

$$\exp\left(-\frac{2\alpha}{(h_n^{\mathbf{a}})^r}\right) = \exp\left(-2\alpha\sqrt{\frac{\log n}{2\beta}} + \frac{\alpha A}{\beta}\right)(1 + o(1)),$$

while for the variance term exponent

$$\frac{1}{n}\exp\left(-\frac{2\beta}{(h_n^{\mathbf{a}})^s}\right) = \exp\left(-2A\sqrt{\frac{\log n}{2\beta}}\right).$$

Since $A > \alpha$, the bias term of $f_n^{\mathbf{a}}$ asymptotically dominates its variance term. Inequalities (21) and (22) now follow from these remarks and the expressions for φ_n^2, $\varphi_n^2(\mathbb{L}_2)$ in (18) and (19) with $r = s/2$. □

4 Minimax lower bounds, sharp optimality and superefficiency

In this section we establish lower bounds for the risks showing that, under mild additional assumptions, the upper bounds of the previous section cannot be improved (in a minimax sense on the class of densities $\mathcal{A}_{\alpha,r}(L)$) not only among kernel estimators, but also among all estimators. In other words, the estimators suggested in the previous section attain optimal rates of convergence on $\mathcal{A}_{\alpha,r}(L)$ with optimal exact constants.

We suppose that the following assumption holds.

Assumption (ND). *There exist constants $u_1 > 0$, $B > 0$ and $\gamma_1 \in \mathbb{R}$ such that $\Phi^{\varepsilon}(u)$ is twice continuously differentiable for $|u| \geq u_1$ with the derivatives satisfying*

$$\max\{|(\Phi^{\varepsilon}(u))'|, |(\Phi^{\varepsilon}(u))''|\} \leq B|u|^{\gamma_1} \exp(-\beta|u|^s),$$

where $\beta > 0$ and $s > 0$ are the same as in Assumption (N).

Note that this assumption is satisfied for the examples of popular noise densities mentioned in the Introduction.

Theorem 4 *Let $\alpha > 0, L > 0, 0 < r < s \leq 2$, and suppose that Assumption (ND) and the right hand inequality in (1) hold. Then*

$$\liminf_{n\to\infty} \inf_{T_n} R_n(x, T_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} \geq 1, \qquad \forall\, x \in \mathbb{R}, \tag{23}$$

and

$$\liminf_{n\to\infty} \inf_{T_n} R_n(\mathbb{L}_2, T_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) \geq 1, \tag{24}$$

where $\inf_{T_n}$ denotes the infimum over all estimators and the rates φ_n, $\varphi_n(\mathbb{L}_2)$ are defined in (14) and (15).

Proof of Theorem 4 is given in Section 5.

Theorems 1,2 and 4 immediately imply the following result on sharp asymptotic minimaxity of the estimators constructed in Section 3.

Theorem 5 *Let $\alpha > 0, L > 0, 0 < r < s \leq 2$, let Assumptions (N), (ND) hold and $\Phi^{\varepsilon}(u) \neq 0, \forall\, u \in \mathbb{R}$. Then the kernel estimator $\hat{f}_n$ with bandwidth defined by (13) (or with bandwidth defined by (20) if $r < s/2$) is sharp asymptotically minimax on $\mathcal{A}_{\alpha,r}(L)$ both in pointwise and in $\mathbb{L}_2$ sense:*

$$\lim_{n\to\infty} R_n(x, \hat{f}_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} = \lim_{n\to\infty} \inf_{T_n} R_n(x, T_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} = 1, \qquad \forall\, x \in \mathbb{R}, \tag{25}$$

$$\lim_{n\to\infty} R_n(\mathbb{L}_2, \hat{f}_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) = \lim_{n\to\infty} \inf_{T_n} R_n(\mathbb{L}_2, T_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2}(\mathbb{L}_2) = 1. \tag{26}$$

This is the main result of the paper. It shows that the kernel estimator $\hat{f}_n$ with a properly chosen bandwidth h_n is sharp optimal in asymptotically minimax sense on $\mathcal{A}_{\alpha,r}(L)$ and that for $r < s/2$ the estimator $f_n^{\mathbf{a}}$ is sharp adaptive in asymptotically minimax sense on $\mathcal{A}_{\alpha,r}(L)$. Sharp adaptation is thus obtained by direct tuning of the smoothing parameter without any additional adaptation rule. This is one of the effects of dominating bias. Theorem 5 also provides exact asymptotical expressions for minimax risks on $\mathcal{A}_{\alpha,r}(L)$ under the pointwise and the $\mathbb{L}_2$ losses: it states that they are equal to φ_n^2 and $\varphi_n^2(\mathbb{L}_2)$ respectively.

Thus, φ_n^2 and $\varphi_n^2(\mathbb{L}_2)$ can be chosen as reference values to determine efficiency of estimators. An interesting question is whether there exist superefficient estimators $\tilde{f}_n$, i.e. such that

$$\sup_{x\in\mathbb{R}} E_f\left[|\tilde{f}_n(x) - f(x)|^2\right] = o(\varphi_n^2) \quad \text{and} \quad E_f\left[\|\tilde{f}_n - f\|_2^2\right] = o(\varphi_n^2(\mathbb{L}_2)), \tag{27}$$

as $n \to \infty$, for any fixed $f \in \mathcal{A}_{\alpha,r}(L)$. The answer to this question is positive, as shows the next proposition.

Proposition 3 *Let the conditions of Theorem 1 hold. Let $\tilde{f}_n$ be the kernel estimator $\hat{f}_n$ with bandwidth defined by (13) (or by (20) if $r < s/2$). Then $\tilde{f}_n$ satisfies (27). If, moreover, the conditions of Theorem 5 hold, $\tilde{f}_n$ is superefficient in the sense that*

$$\lim_{n\to\infty} \frac{E_f[|\tilde{f}_n(x) - f(x)|^2]}{\inf_{T_n}\sup_{f\in\mathcal{A}_{\alpha,r}(L)} E_f\left[|T_n(x) - f(x)|^2\right]} = 0, \quad \forall x \in \mathbb{R}, \tag{28}$$

$$\lim_{n\to\infty} \frac{E_f[\|\tilde{f}_n - f\|_2^2]}{\inf_{T_n}\sup_{f\in\mathcal{A}_{\alpha,r}(L)} E_f\left[\|T_n - f\|_2^2\right]} = 0. \tag{29}$$

Proof. Consider the kernel estimator $\hat{f}_n$ with bandwidth defined by (13). Instead of using Proposition 1 to bound the bias term, we apply directly (4) for the pointwise risk and (5) for the $\mathbb{L}_2$-risk which yields that, for any fixed $f \in \mathcal{A}_{\alpha,r}(L)$,

$$\begin{aligned}\sup_{x\in\mathbb{R}} |E_f\hat{f}_n(x) - f(x)|^2 &= o\left(h_*^{r-1}\exp(-2\alpha/h_*^r)\right) = o(\varphi_n^2),\\ \|E_f\hat{f}_n - f\|_2^2 &= o\left(\exp(-2\alpha/h_*^r)\right) = o(\varphi_n^2(\mathbb{L}_2)),\end{aligned}$$

as $n \to \infty$. Now, Proposition 2 and (68) of Lemma 8 imply that the variance terms are also $o(\varphi_n^2)$ and $o(\varphi_n^2(\mathbb{L}_2))$, as $n \to \infty$, respectively. Hence, (27) follows and implies (28) and (29), in view of Theorem 5. The case where the bandwidth is defined by (20) and $r < s/2$ is treated similarly. □

The result of Proposition 3 is explained by the fact that the value of the minimax risk in the denominator of (29) is attained (up to a $1 + o(1)$ factor) on the densities that depend on n, while in the numerator we have a fixed density f. Such a superefficiency property occurs in other nonparametric problems (see e.g. Brown, Low and Zhao (1997) or Tsybakov (2004), Chapter 3), where it is proved for various adaptive estimators. On the contrary, non-adaptive asymptotically minimax estimators, for example, the Pinsker estimator which is efficient for ellipsoids in gaussian sequence model, are not superefficient and turn out to be inadmissible (Tsybakov (2004), Section 3.8). Compared with that, the result of Proposition 3 is somewhat surprising, because it states that a non-adaptive asymptotically minimax estimator $\hat{f}_n$ with bandwidth defined by (13) is superefficient. This provides a simple counter-example of a superefficient nonparametric estimator which is not adaptive. We conjecture that this is a general property of nonparametric problems with dominating bias.

5 Proof of Theorem 4

5.1 General scheme of the proof

We use the method of proving lower bounds by reduction to the problem of testing two simple hypotheses (cf. e.g. Tsybakov (2004), Chapter 2). Namely, we define two properly chosen probability densities f_{n1} and f_{n2}, depending on n and belonging to $\mathcal{A}_{\alpha,r}(L)$ and we bound the minimax risk as follows

$$\begin{aligned} \inf_{T_n} R_n(T_n, \mathcal{A}_{\alpha,r})\psi_n^{-2} &\geq \inf_{T_n} \max_{f \in \{f_{n1}, f_{n2}\}} E_f d^2(T_n, f)\psi_n^{-2} \\ &\geq \inf_{T_n} \max_{f \in \{f_{n1}, f_{n2}\}} \left(E_f d(T_n, f)\right)^2 \psi_n^{-2}, \end{aligned} \tag{30}$$

where $R_n(T_n, \mathcal{A}_{\alpha,r}(L))$ is either $R_n(x, T_n, \mathcal{A}_{\alpha,r}(L))$ or $R_n(\mathbb{L}_2, T_n, \mathcal{A}_{\alpha,r}(L))$, ψ_n is defined as φ_n or $\varphi_n(\mathbb{L}_2)$ (cf. (14) and (15)) respectively and $d(T_n, f)$ stands for the distance $|T_n(x) - f(x)|$ at a fixed point x or the $\mathbb{L}_2$-distance $\|T_n - f\|_2$ respectively. Hence, to prove the theorem it remains to show that

$$R \stackrel{\text{def}}{=} \inf_{T_n} \max_{f \in \{f_{n1}, f_{n2}\}} E_f d(T_n, f) \geq \psi_n(1 + o(1)), \tag{31}$$

as $n \to \infty$, for both pointwise and $\mathbb{L}_2$ distances $d(\cdot, \cdot)$. This will be done by application of Lemma 4 of the Appendix. According to Lemma 4, (31) is satisfied if the functions f_{n1} and f_{n2} are chosen such that

$$d(f_{n1}, f_{n2}) \geq 2\psi_n(1 + o(1)), \text{ as } n \to \infty, \tag{32}$$

$$\chi^2(P_{f_{n1}}, P_{f_{n2}}) = o(1), \text{ as } n \to \infty, \tag{33}$$

where $\chi^2(P_{f_{n1}}, P_{f_{n2}})$ is the χ^2-divergence between the probability measures $P_{f_{n1}}$ and $P_{f_{n2}}$ (recall that P_f denotes the joint distribution of $Y_1, \ldots, Y_n$ when the underlying probability density of X_i's is f). Thus, to prove Theorem 4 it suffices to construct two functions f_{n1} and f_{n2} belonging to $\mathcal{A}_{\alpha,r}(L)$ and satisfying (32) − (33). Since $P_{f_{nj}}$ is a product of n identical probability measures corresponding to the density $f^Y_{nj} = f_{nj} * f^\varepsilon$, for $j = 1, 2$, we have $\chi^2(P_{f_{n1}}, P_{f_{n2}}) \leq Cn\chi^2(f^Y_{n1}, f^Y_{n2})$ if $\chi^2(f^Y_{n1}, f^Y_{n2}) \leq 1/n$, where C is a finite constant and

$$\chi^2(f^Y_{n1}, f^Y_{n2}) = \int \frac{(f^Y_{n1} - f^Y_{n2})^2}{f^Y_{n1}}(x)dx$$

(cf. e.g. Tsybakov (2004), p. 72). Therefore, (33) follows from

$$n\chi^2(f^Y_{n1}, f^Y_{n2}) \to 0, \text{ as } n \to \infty. \tag{34}$$

We now proceed to the construction of densities f_{n1}, $f_{n2} \in \mathcal{A}_{\alpha,r}(L)$ satisfying (34) and (32) for pointwise and $\mathbb{L}_2$-distances $d(\cdot,\cdot)$.

Consider a density f_0 of a symmetric stable law whose characteristic function is

$$\Phi_0(u) = \begin{cases} \exp\left(-|c_0 u|^r\right), & \text{if } 1 < r < 2, \\ \exp\left(-|c_0 u|\right), & \text{if } 0 < r \le 1, \end{cases}$$

where $c_0 > \max\{\alpha^{1/r}, \alpha\}$. Clearly, for any $0 < a < 1$ there exists $c_0 > 0$ large enough so that $f_0 \in \mathcal{A}_{\alpha,r}(a^2 L)$. In view of Lemma 7, there exists $c_1' > 0$ such that

$$f_0(x) = \frac{1}{c_0} p\left(\frac{x}{c_0}\right) \ge \frac{c_1'}{|x|^{\max\{r+1,2\}} + 1}, \tag{35}$$

for all $x \in \mathbb{R}$, where p is the density of stable symmetric distribution with characteristic function $\exp(-|t|^{\max\{r,1\}})$, $0 < r < 2$. Let $h_+ = h_+(n)$ be the unique solution of the equation

$$\frac{2\alpha}{h_+^r} + \frac{2\beta}{h_+^s} = \log n + (\log\log n)^2. \tag{36}$$

Note that h_+ is analogous to h_* defined by (13) with the only difference that the $(\log\log n)^2$ term changes the sign.

We define the densities f_{n1} and f_{n2} by their characteristic functions

$$\Phi_{n1}(u) = \Phi_0(u) + \Phi^H(u, h_+), \quad \Phi_{n2}(u) = \Phi_0(u) - \Phi^H(u, h_+), \quad u \in \mathbb{R}, \tag{37}$$

where $u \mapsto \Phi^H(u,h)$ with $h > 0$ will be called *perturbation function* and will be defined differently for the pointwise distance and the $\mathbb{L}_2$-distance. The construction of perturbation functions will be based on the following lemma.

Lemma 1 *For any $\delta > 0$ and any $D > 4\delta$ there exists a function $\Phi^G : \mathbb{R} \to [0,1]$ such that*

(i) Φ^G is 3 times continuously differentiable on $\mathbb{R}$ and the first 3 derivatives of Φ^G are uniformly bounded on $\mathbb{R}$,

(ii) Φ^G is compactly supported on $(\delta, D-\delta)$ and

$$I(2\delta \le u \le D - 2\delta) \le \Phi^G(u) \le I(\delta \le u \le D - \delta),$$

for all $u \in \mathbb{R}$.

Proof of Lemma 1. Denote by J_0 the 5-fold convolution of the indicator function $I(|u| \leq 1)$ with itself. Let $J : \mathbb{R} \to [0, \infty)$ be a rescaling of J_0 such that the support of J is $(-1, 1)$ and $\int J(x)dx = 1$. Then J_0 and J are 3 times continuously differentiable on $\mathbb{R}$. For $\delta > 0$ and $D > 4\delta$ define

$$\Phi^G(u) = \int_{u-D+3\delta/2}^{u-3\delta/2} \frac{2}{\delta} J\left(\frac{2x}{\delta}\right) dx.$$

Clearly, Φ^G is 3 times continuously differentiable on $\mathbb{R}$ and $0 \leq \Phi^G(u) \leq 1$, $\forall u \in \mathbb{R}$. Moreover, *supp* $\Phi^G = (\delta, D - \delta)$ and for any $u \in (2\delta, D - 2\delta)$ we have $\Phi^G(u) = \int_{-1}^{1} J(x)dx = 1$. □

5.2 Lower bound at a fixed point

Without loss of generality, we will prove the lower bound for the distance $d(f, g) = |f(0) - g(0)|$ at the point $x = 0$ (if $x \neq 0$ it suffices to shift the functions f_{n1} and f_{n2} at x). Define the perturbation function

$$\Phi^H(u, h) = \sqrt{2\pi\alpha r L}\, h^{(1-r)/2} \exp\left(\frac{\alpha}{h^r}\right) \exp\left(-2\alpha\, |u|^r\right) \Phi^G\left(|u|^r - \frac{1}{h^r}\right), \qquad (38)$$

where Φ^G is a function satisfying the properties given in Lemma 1 for some $\delta > 0$ and $D > 4\delta$.

Most of the computations below work when Φ^G is replaced by an indicator function of the interval $[0, D]$. However, we obviously need a continuous perturbation function Φ^H that satisfies $\Phi^H(0) = 0$ to ensure that f_{n1} and f_{n2} integrate to 1 and that is smooth enough to allow an appropriate bound on the χ^2-divergence.

Lemma 2 *Let f_{n1} and f_{n2} be the functions defined by their Fourier transforms* (37), (38) *with Φ^G satisfying the properties given in Lemma* 1. *Then we have the following.*

1. *The functions f_{n1} and f_{n2} are probability densities for any n large enough.*

2. *The functions f_{n1} and f_{n2} belong to $\mathcal{A}_{\alpha,r}(L)$ for n large enough if $c_0 > 0$ in the definition of f_0 large enough.*

3. *The distance between f_{n1} and f_{n2} at $x = 0$ satisfies*

$$|f_{n1}(0) - f_{n2}(0)| \geq 2\varphi_n[e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}](1 + o(1)),$$

as $n \to \infty$.

4. *The χ^2-divergence $\chi^2\left(f_{n1}^Y, f_{n2}^Y\right)$ satisfies* (34).

Proof. 1. Clearly, $\Phi^H(\cdot, h)$ is an even, 3 times continuously differentiable function on $\mathbb{R}$ having a compact support. It is easy to see that the integrals $\int |\Phi^H(u,h)|du$ and $\int |\partial^3\Phi^H(u,h)/\partial u^3|du$ are bounded uniformly over $0 < h \le h_0$ for any $h_0 > 0$. Integration by parts yields that the inverse Fourier transform of $\Phi^H(\cdot, h)$ can be written as

$$H(x,h) \stackrel{\text{def}}{=} \frac{1}{2\pi}\int \cos(xu)\Phi^H(u,h)du = -\frac{1}{2\pi x^3}\int \sin(xu)\frac{\partial^3\Phi^H(u,h)}{\partial u^3}du \tag{39}$$

for all $x \in \mathbb{R}$ and $0 < h \le h_0$. Thus, there exists a constant $C_H < \infty$ independent of n and such that

$$|H(x,h_+)| \le C_H(|x|^3+1)^{-1}, \text{ for all } x \in \mathbb{R}. \tag{40}$$

Denote by Dom the common support of the functions $\Phi^G(|u|^r - 1/h_+^r)$ and $\Phi^H(u,h_+)$:

$$Dom \stackrel{\text{def}}{=} \left\{u : |u|^r - \frac{1}{h_+^r} \in [\delta, D-\delta]\right\} = \left\{u : \left(\delta + \frac{1}{h_+^r}\right)^{1/r} \le |u| \le \left(D - \delta + \frac{1}{h_+^r}\right)^{1/r}\right\}.$$

Using the fact that $\left(\delta + 1/h_+^r\right)^{1/r} \to \infty$, as $n \to \infty$, for any fixed $\delta > 0$ and applying (63) of Lemma 6 in the Appendix, we find

$$\begin{aligned}
\|H(\cdot,h_+)\|_\infty &\stackrel{\text{def}}{=} \sup_{x\in\mathbb{R}} |H(x,h_+)| \le \frac{1}{2\pi}\int |\Phi^H(u,h_+)|du \\
&\le \sqrt{\frac{\alpha r L}{2\pi}} h_+^{(1-r)/2} \exp\left(\alpha/h_+^r\right) \int_{Dom} \exp(-2\alpha|u|^r)du \\
&\le c h_+^{(r-1)/2} \exp(-\alpha/h_+^r) = o(1), \text{ as } n \to \infty,
\end{aligned} \tag{41}$$

where $c > 0$ is a finite constant.

Now, $f_{n1}(x) = f_0(x) + H(x,h_+)$, $f_{n2}(x) = f_0(x) - H(x,h_+)$. Choose $A > 0$ large enough so that for $|x| > A$ we have $C_H(|x|^3+1)^{-1} < c_1'(|x|^{\max\{r+1,2\}}+1)^{-1}$ (note that $\max\{r+1, 2\} < 3$). Then, in view of (35) and (40), $f_{nj}(x) > 0$, $j = 1, 2$, for $|x| > A$. Now, if n is large enough, $f_{nj}(x) > 0$ also for $|x| \le A$ since $\inf_{|x|\le A} f_0(x) > 0$ (cf. (35)) and (41) holds.

Thus, $f_{nj}(x) > 0$, $j = 1, 2$, for all $x \in \mathbb{R}$ if n is large enough. It remains to note that f_{n1} and f_{n2} integrate to 1 since $\int H(x,h_+)dx = \Phi^H(0,h_+) = 0$ (indeed, $0 \notin supp\ \Phi^H(\cdot,h_+) = Dom$).

2. We have, by (38) and Lemma 1,

$$\begin{aligned}
&\int \left|\Phi^H(u,h_+)\right|^2 \exp(2\alpha|u|^r)\,du \\
\le\ & 2\pi\alpha r L h_+^{1-r} \exp\left(\frac{2\alpha}{h_+^r}\right) \int_{Dom} \exp(-2\alpha|u|^r)\,du \\
\le\ & 4\pi\alpha r L h_+^{1-r} \exp\left(\frac{2\alpha}{h_+^r}\right) \int_{(\delta+1/h_+^r)^{1/r}}^{\infty} \exp(-2\alpha u^r)du.
\end{aligned}$$

By Lemma 6,

$$\begin{aligned}\int_{(\delta+1/h_+^r)^{1/r}}^{\infty} \exp(-2\alpha u^r)du &= \frac{1}{2\alpha r}\left(\delta+\frac{1}{h_+^r}\right)^{(1-r)/r}\exp\left(-2\alpha\left(\delta+\frac{1}{h_+^r}\right)\right)(1+o(1)) \\ &= \frac{h_+^{r-1}}{2\alpha r}\exp\left(-\frac{2\alpha}{h_+^r}\right)\exp(-2\alpha\delta)(1+\delta h_+^r)^{(1-r)/r}(1+o(1)),\end{aligned}$$

as $n\to\infty$. We get therefore,

$$\int |\Phi^H(x,h_+)|^2\exp(2\alpha|u|^r)du \le 2\pi L\exp(-2\alpha\delta)(1+o(1)), \tag{42}$$

as $n\to\infty$, for any fixed $\delta>0$. Now, choose $c_0>0$ in the definition of f_0 large enough to guarantee that $f_0\in\mathcal{A}_{\alpha,r}(a^2L)$ with $a=1-e^{-\alpha\delta/2}$. This and (42) imply

$$\begin{aligned}\left(\int |\Phi_{nj}(u)|^2\exp(2\alpha|u|^r)du\right)^{1/2} &\le \|\Phi_0(\cdot)\exp(\alpha|\cdot|^r)\|_2+\|\Phi^H(\cdot,h_+)\exp(\alpha|\cdot|^r)\|_2 \\ &\le (1-e^{-\alpha\delta/2})\sqrt{2\pi L}+e^{-\alpha\delta}\sqrt{2\pi L}(1+o(1)) \\ &\le \sqrt{2\pi L}, \quad j=1,2,\end{aligned}$$

for n large enough and any fixed $\delta>0$.

3. Using the left inequality in (ii) of Lemma 1 we get

$$\begin{aligned}|f_{n1}(0)-f_{n2}(0)|^2 &= \frac{1}{(2\pi)^2}\left|\int(\Phi_{n1}(u)-\Phi_{n2}(u))\,du\right|^2=\frac{4}{(2\pi)^2}\left|\int\Phi^H(u,h_+)\,du\right|^2 \\ &= \frac{2\alpha rLh_+^{1-r}}{\pi}\exp\left(\frac{2\alpha}{h_+^r}\right)\left|\int\exp(-2\alpha|u|^r)\,\Phi^G\left(|u|^r-\frac{1}{h_+^r}\right)du\right|^2 \\ &\ge \frac{2\alpha rLh_+^{1-r}}{\pi}\exp\left(\frac{2\alpha}{h_+^r}\right)\left|2\int_{(2\delta+1/h_+^r)^{1/r}}^{(D-2\delta+1/h_+^r)^{1/r}}\exp(-2\alpha u^r)\,du\right|^2. \end{aligned} \tag{43}$$

By (63) of Lemma 6 in the Appendix,

$$\begin{aligned}&\int_{(2\delta+1/h_+^r)^{1/r}}^{(D-2\delta+1/h_+^r)^{1/r}}\exp(-2\alpha u^r)\,du \\ &= \frac{h_+^{r-1}}{2\alpha r}\exp\left(-\frac{2\alpha}{h_+^r}\right)\left[(1+2\delta h_+^r)^{(1-r)/r}e^{-4\alpha\delta}(1+o(1))\right. \\ &\quad \left.-(1+(D-2\delta)h_+^r)^{(1-r)/r}e^{-2\alpha(D-2\delta)}(1+o(1))\right] \\ &= \frac{h_+^{r-1}}{2\alpha r}\exp\left(-\frac{2\alpha}{h_+^r}\right)[e^{-4\alpha\delta}-e^{-2\alpha(D-2\delta)}](1+o(1)),\end{aligned} \tag{44}$$

as $n\to\infty$. The expression in square brackets here is positive since $D>4\delta$. Combining (43) and (44) and using (74) of Lemma 9 in the Appendix together with

(14) we get

$$
\begin{aligned}
|f_{n1}(0) - f_{n2}(0)|^2 &\geq 4\left[\frac{L}{2\pi\alpha r}h_+^{r-1}\exp\left(-\frac{2\alpha}{h_+^r}\right)\right][e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}]^2(1+o(1)) \\
&= 4\left[\frac{L}{2\pi\alpha r}h_*^{r-1}\exp\left(-\frac{2\alpha}{h_*^r}\right)\right][e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}]^2(1+o(1)) \\
&= 4\varphi_n^2[e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}]^2(1+o(1)),
\end{aligned}
$$

as $n \to \infty$.

4. Inequalities (35), (40), (41) and the fact that $r < 2$ imply the existence of a constant $c_2' > 0$ independent of n and such that

$$
f_{n1}(x) \geq \frac{c_2'}{|x|^{\max\{r+1,2\}} + 1}, \quad \forall x \in \mathbb{R},
$$

for all n large enough. Since f^ε is a probability density, we have $\int_{-M}^{M} f^\varepsilon(x)dx \geq 1/2$ for a constant $M > 1$ large enough. Hence,

$$
\begin{aligned}
f_{n1}^Y(x) &\geq \int_{-M}^{M} f_{n1}(x-y)f^\varepsilon(y)dy \geq \frac{c_2'}{2}\inf_{|y|\leq M}\left[\frac{1}{|x-y|^{\max\{r+1,2\}}+1}\right] \\
&\geq c_3'\min\left\{\frac{1}{M^{\max\{r+1,2\}}}, \frac{1}{|x|^{\max\{r+1,2\}}}\right\} \qquad (45)
\end{aligned}
$$

where n and M are large enough, $c_3' > 0$ is independent of n, and the last inequality is obtained by considering separately $|x| \leq M$ and $|x| > M$. Thus

$$
\begin{aligned}
n\chi^2(f_{n1}^Y, f_{n2}^Y) &= n\int \frac{(f_{n2}^Y - f_{n1}^Y)^2(x)}{f_{n1}^Y(x)}dx = 4n\int \frac{(H * f^\varepsilon)^2(x)}{f_{n1}^Y(x)}dx \\
&\leq \frac{4}{c_3'}\Bigg(nM^{\max\{r+1,2\}}\int_{|x|\leq M}(H * f^\varepsilon)^2(x)dx \\
&\quad + n\int_{|x|>M}|x|^{\max\{r+1,2\}}(H * f^\varepsilon)^2(x)dx\Bigg) \\
&\leq (4M^3/c_3')(T_{n1} + T_{n2}), \qquad (46)
\end{aligned}
$$

for n and M large enough, where $H(x) = H(x, h_+)$ for brevity and

$$
T_{n1} = n\|H * f^\varepsilon\|_2^2, \quad T_{n2} = n\int |x|^4(H * f^\varepsilon)^2(x)dx. \qquad (47)
$$

Using Plancherel's formula and the right hand inequality in (1) we get, for n large enough,

$$\begin{aligned}
\|H * f^{\varepsilon}\|_2^2 &= \frac{1}{2\pi}\int \left|\Phi^H(u,h_+)\Phi^{\varepsilon}(u)\right|^2 du \\
&\le b_{\max}^2 \alpha r L h_+^{1-r}\exp\left(\frac{2\alpha}{h_+^r}\right)\int_{Dom}|u|^{2\gamma'}\exp(-4\alpha|u|^r - 2\beta|u|^s)du \\
&\le 2b_{\max}^2 \alpha r L h_+^{1-r}\exp\left(\frac{2\alpha}{h_+^r}\right)\int_{(\delta+1/h_+^r)^{1/r}}^{\infty} u^{2\gamma'}\exp(-4\alpha u^r - 2\beta u^s)du \\
&\le 2b_{\max}^2 \alpha r L h_+^{1-r}\exp\left(-\frac{2\alpha}{h_+^r}\right)\int_{1/h_+}^{\infty} u^{2\gamma'}\exp(-2\beta u^s)du. \qquad (48)
\end{aligned}$$

The last integral is evaluated using (63) of Lemma 6 in the Appendix:

$$\int_{1/h_+}^{\infty} u^{2\gamma'}\exp(-2\beta u^s)du = \frac{h_+^{s-2\gamma'-1}}{2\beta s}\exp\left(-\frac{2\beta}{h_+^s}\right)(1+o(1)), \qquad (49)$$

as $n \to \infty$. This, together with (48) and (75) of Lemma 9 in the Appendix, yields

$$\|H * f^{\varepsilon}\|_2^2 \le C h_+^{s-2\gamma'-r}\exp\left(-\frac{2\alpha}{h_+^r} - \frac{2\beta}{h_+^s}\right) = o\left(\frac{1}{n}\right), \qquad (50)$$

as $n \to \infty$, where $C > 0$ is a constant. Thus,

$$T_{n1} = o(1), \text{ as } n \to \infty. \qquad (51)$$

Now, assume that n is large enough to have $(\delta + 1/h_+^r)^{1/r} > \max(u_0, u_1)$, where $u_0 > 0$, $u_1 > 0$ are the constants in Assumptions (N) and (ND). Then $\Phi^G(|u|^r - 1/h_+^r) = 0$ for $|u| \le \max(u_0, u_1)$, and thus the function $\Phi^H(\cdot, h_+)\Phi^{\varepsilon}(\cdot)$ is twice continuously differentiable on $\mathbb{R}$. Using Assumption (ND), the right hand inequality in (1) and the fact that Φ^G, together with its first two derivatives, is uniformly bounded on $\mathbb{R}$ we find that there exist constants $B_1 < \infty$ and $a \in \mathbb{R}$ such that, for n large enough and all $u \in \mathbb{R}$,

$$\left|(\Phi^H(u,h_+)\Phi^{\varepsilon}(u))''\right| \le B_1 h_+^{(1-r)/2}\exp\left(\frac{\alpha}{h_+^r}\right)|u|^a\exp(-2\alpha|u|^r - \beta|u|^s). \qquad (52)$$

Thus, for n large enough, we have, by Plancherel's formula for derivatives and (52),

$$\begin{aligned}
T_{n2} &= \frac{n}{2\pi}\int \left|(\Phi^H(u,h_+)\Phi^{\varepsilon}(u))''\right|^2 du \\
&\le \frac{n}{2\pi}B_1^2 h_+^{1-r}\exp\left(\frac{2\alpha}{h_+^r}\right)\int_{Dom}|u|^{2a}\exp(-4\alpha|u|^r - 2\beta|u|^s)du \\
&\le \frac{n}{\pi}B_1^2 h_+^{1-r}\exp\left(\frac{2\alpha}{h_+^r}\right)\int_{(\delta+1/h_+^r)^{1/r}}^{\infty} u^{2a}\exp(-4\alpha u^r - 2\beta u^s)du \\
&\le \frac{n}{\pi}B_1^2 h_+^{1-r}\exp\left(-\frac{2\alpha}{h_+^r}\right)\int_{1/h_+}^{\infty} u^{2a}\exp(-2\beta u^s)du. \qquad (53)
\end{aligned}$$

Plugging (49) with $\gamma' = a$ into (53) and using (75) of Lemma 9 in the Appendix we get

$$T_{n2} \le Cnh_+^{-2a+s-r} \exp\left(-\frac{2\alpha}{h_+^r} - \frac{2\beta}{h_+^s}\right)(1+o(1)) = o(1), \tag{54}$$

as $n \to \infty$, where $C > 0$ is a constant.

Combining (46), (51) and (54) we get that $n\chi^2(f_{n1}^Y, f_{n2}^Y) \to 0$, as $n \to \infty$. □

Proof of (23)**.** We use the general scheme of Section 5.1 with $d(f_{n1}, f_{n2}) = |f_{n1}(0) - f_{n2}(0)|$. Choose $c_0 > 0$ in the definition of f_0 large enough to guarantee that assertion 2 of Lemma 2 holds. Lemma 2 implies that (34) and thus (33) are satisfied and that (32) holds with

$$\psi_n = \varphi_n[e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}].$$

Therefore, Lemma 4 of the Appendix implies that

$$R \ge \varphi_n[e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}](1+o(1)),$$

as $n \to \infty$, where R is defined in (31). This and (30) yield that, as $n \to \infty$,

$$\inf_{T_n} R_n(0, T_n, \mathcal{A}_{\alpha,r}(L))\varphi_n^{-2} \ge [e^{-4\alpha\delta} - e^{-2\alpha(D-2\delta)}](1+o(1)).$$

Taking limits as $n \to \infty$ and then as $D \to \infty$ and $\delta \to 0$ we get (23) for $x = 0$. The proof for $x \ne 0$ is analogous (see the remark at the beginning of this section). □

5.3 Lower bound in $\mathbb{L}_2$

Introduce the perturbation function

$$\Phi^H(u,h) = \sqrt{2\pi\alpha r L(d-1)}\; h^{(1-r)/2} e^{(d-1)\alpha/h^r} \exp\left(-\alpha d|u|^r\right)\Phi^G\left(|u|^r - \frac{1}{h^r}\right), \tag{55}$$

where Φ^G is a function satisfying the properties given in Lemma 1 and $d = d(\delta) > 1$ is a constant depending on the value δ that appears in the construction of Φ^G. The argument below is similar to that of Section 5.2, modulo the choice of the perturbation function (55) which is slightly different from (38). The argument goes through with d such that $d(\delta) \to \infty$ and $\delta d(\delta) \to 0$ as $\delta \to 0$, but we will set for simplicity $d(\delta) = \delta^{-1/2}$ and assume that $0 < \delta < 1$, which ensures that $d(\delta) > 1$.

Lemma 3 *Let f_{n1} and f_{n2} be the functions defined by their Fourier transforms* (37), (55) *with Φ^G satisfying the properties of Lemma* 1 *and $0 < \delta < 1$. Then we have the following.*

1. *The functions f_{n1} and f_{n2} are probability densities for n large enough.*
2. *The functions f_{n1} and f_{n2} belong to $\mathcal{A}_{\alpha,r}(L)$ for n large enough if $c_0 > 0$ in the definition of f_0 large enough.*

3. *The* $\mathbb{L}_2$ *distance between* f_{n1} *and* f_{n2} *satisfies*

$$\|f_{n1} - f_{n2}\|_2 \geq 2\varphi_n(\mathbb{L}_2)\left((1-\sqrt{\delta})[e^{-4\alpha\sqrt{\delta}} - e^{-2\alpha(D-2\delta)/\sqrt{\delta}}]\right)^{1/2}(1+o(1)),$$

as $n \to \infty$.

4. *The* χ^2*-divergence* $\chi^2\left(f^Y_{n1}, f^Y_{n2}\right)$ *satisfies* (34).

Proof. 1. The argument is analogous to the proof of assertion 1 of Lemma 2. In particular, one also has $|H(x,h)| \leq C'_H(|x|^3+1)^{-1}$, $\forall x \in \mathbb{R}$, and $\|H(\cdot, h_+)\|_\infty = o(1)$, as $n \to \infty$, for some constant $C'_H < \infty$. We omit the details.

2. We have by (37) and Lemma 1

$$\begin{aligned}
&\int \left|\Phi^H(u,h_+)\right|^2 \exp\left(2\alpha\left|u\right|^r\right) du \\
\leq\ & 2\pi\alpha r L(d-1)h_+^{1-r}\exp\left(\frac{2(d-1)\alpha}{h_+^r}\right)\int_{Dom}\exp\left(-2\alpha(d-1)\left|u\right|^r\right) du \\
\leq\ & 4\pi\alpha r L(d-1)h_+^{1-r}\exp\left(\frac{2(d-1)\alpha}{h_+^r}\right)\int_{(\delta+1/h_+^r)^{1/r}}^{\infty}\exp\left(-2\alpha(d-1)u^r\right) du.
\end{aligned}$$

By Lemma 6,

$$\begin{aligned}
&\int_{(\delta+1/h_+^r)^{1/r}}^{\infty}\exp\left(-2\alpha(d-1)u^r\right) du \\
=\ & \frac{1}{2\alpha(d-1)r}\left(\delta+\frac{1}{h_+^r}\right)^{(1-r)/r}\exp\left(-2\alpha(d-1)\left(\delta+\frac{1}{h_+^r}\right)\right)(1+o(1)) \\
=\ & \frac{h_+^{r-1}}{2\alpha(d-1)r}\exp\left(-\frac{2(d-1)\alpha}{h_+^r}\right)\exp\left(-2\alpha(d-1)\delta\right)\left(1+\delta h_+^r\right)^{(1-r)/r}(1+o(1)),
\end{aligned}$$

as $n \to \infty$. We get therefore,

$$\int |\Phi^H(u,h_+)|^2 \exp(2\alpha|u|^r)du \leq 2\pi L\exp(-2\alpha(d-1)\delta)(1+o(1)),$$

as $n \to \infty$, for any fixed $\delta > 0$. Now, since $d = \delta^{-1/2}$, we get that the last exponent is strictly less than 1 for $0 < \delta < 1$, and thus the argument similar to that after formula (42) can be applied to show that

$$\int |\Phi_{nj}(u)|^2 \exp(2\alpha|u|^r)du \leq 2\pi L, \quad j = 1,2,$$

for n large enough, if $c_0 > 0$ in the definition of f_0 is chosen large enough.

3. The $\mathbb{L}_2$ distance is

$$\|f_{n1} - f_{n2}\|_2^2 = \frac{1}{2\pi}\int \left(\Phi_{n1}(u) - \Phi_{n2}(u)\right)^2 du = \frac{4}{2\pi}\int \left|\Phi^H(u, h_+)\right|^2 du$$

$$= \quad 4L\alpha r(d-1)h_+^{1-r}\exp\left(\frac{2(d-1)\alpha}{h_+^r}\right)\int \exp\left(-2\alpha d\,|u|^r\right)\left|\Phi^G\left(|u|^r - \frac{1}{h_+^r}\right)\right|^2 du$$

$$\geq \quad 4L\alpha r(d-1)h_+^{1-r}\exp\left(\frac{2(d-1)\alpha}{h_+^r}\right)\left[2\int_{(2\delta+1/h_+^r)^{1/r}}^{(D-2\delta+1/h_+^r)^{1/r}} \exp\left(-2\alpha d u^r\right) du\right] \quad (56)$$

where we used the left inequality in (ii) of Lemma 2. Lemma 6 implies that (cf. (44)):

$$\int_{(2\delta+1/h_+^r)^{1/r}}^{(D-2\delta+1/h_+^r)^{1/r}} \exp\left(-2\alpha d u^r\right) du$$

$$= \quad \frac{h_+^{r-1}}{2\alpha d r}\exp\left(-\frac{2\alpha d}{h_+^r}\right)[e^{-4\alpha d\delta} - e^{-2\alpha d(D-2\delta)}](1+o(1)),$$

as $n \to \infty$. Substituting this into (56) and using (74) of Lemma 9 we obtain

$$\begin{aligned}\|f_{n1} - f_{n2}\|_2^2 \quad &\geq \quad 4L\frac{d-1}{d}\exp\left(-\frac{2\alpha}{h_+^r}\right)[e^{-4\alpha d\delta} - e^{-2\alpha d(D-2\delta)}](1+o(1)) \\ &= \quad 4L\exp\left(-\frac{2\alpha}{h_*^r}\right)(1-\sqrt{\delta})[e^{-4\alpha\sqrt{\delta}} - e^{-2\alpha(D-2\delta)/\sqrt{\delta}}](1+o(1)) \\ &= \quad 4\varphi_n^2(\mathbb{L}_2)(1-\sqrt{\delta})[e^{-4\alpha\sqrt{\delta}} - e^{-2\alpha(D-2\delta)/\sqrt{\delta}}](1+o(1)),\end{aligned}$$

as $n \to \infty$, (cf. the definition of $\varphi_n(\mathbb{L}_2)$ in (15)).

4. Similarly to the proof of assertion 4 of Lemma 2, we obtain

$$n\chi^2(f_{n1}^Y, f_{n2}^Y) \leq c_4'(T_{n1} + T_{n2}), \quad (57)$$

for n and M large enough, where T_{n1} and T_{n2} are defined in (47) and $c_4' < \infty$ is a constant. The only difference from the proof of Lemma 2 is that the function $H(x) = H(x, h_+)$ is now defined as the inverse Fourier transform of (38) and not as that of (37). As in (48) – (50), we get, for n large enough,

$$\begin{aligned}&T_{n1} = n\|H * f^\varepsilon\|_2^2 \\ \leq \quad & b_{\max}^2 \alpha r L(d-1)n h_+^{1-r}\exp\left(\frac{2(d-1)\alpha}{h_+^r}\right)\int_{Dom} |u|^{2\gamma'}\exp\left(-2\alpha d|u|^r - 2\beta|u|^s\right) du \\ \leq \quad & c' n h_+^{1-r}\exp\left(-\frac{2\alpha}{h_+^r}\right)\int_{1/h_+}^{\infty} u^{2\gamma'}\exp\left(-2\beta u^s\right) du \\ \leq \quad & c'' n h_+^{s-2\gamma'-r}\exp\left(-\frac{2\alpha}{h_+^r} - \frac{2\beta}{h_+^s}\right) = o(1), \quad (58)\end{aligned}$$

as $n \to \infty$, where $c' > 0$ and $c'' > 0$ are some finite constants.

Next, similarly to (52), we have, for n large enough and all $u \in \mathbb{R}$,

$$|(\Phi^H(u,h_+)\Phi^\varepsilon(u))''| \le B_2 h_+^{(1-r)/2} \exp\left(\frac{(d-1)\alpha}{h_+^r}\right) |u|^{a'} \exp(-2\alpha d|u|^r - \beta|u|^s),$$

where $B_2 < \infty$ and $a' \in \mathbb{R}$ are some constants. This implies, as in (53) – (54), that

$$\begin{aligned} T_{n2} &= \frac{n}{2\pi}\int |(\Phi^H(u,h_+)\Phi^\varepsilon(u))''|^2 du \\ &\le \frac{n}{\pi} B_2^2 h_+^{1-r} \exp\left(-\frac{2\alpha}{h_+^r}\right)\int_{1/h_+}^\infty u^{2a'} \exp(-2\beta u^s) du \\ &\le \bar{c} n h_+^{-2a'+s-r} \exp\left(-\frac{2\alpha}{h_+^r} - \frac{2\beta}{h_+^s}\right) = o(1), \end{aligned} \tag{59}$$

as $n \to \infty$, where $\bar{c} > 0$ is finite constant. It remains now to combine (57) – (59).

Proof of (24) is now obtained following the same lines as the proof of (23) in Section 5.2, but with $d(f_{n1}, f_{n2}) = \|f_{n1} - f_{n2}\|_2$ and $\psi_n = \varphi_n(\mathbb{L}_2)\Big((1-\sqrt{\delta})[e^{-4\alpha\sqrt{\delta}} - e^{-2\alpha(D-2\delta)/\sqrt{\delta}}]\Big)^{1/2}$. □

6 Appendix

Let $(\mathcal{X}, \mathcal{A})$ and $(\Theta, \mathcal{T})$ be measurable spaces and let P_1 and P_2 be two probability measures on $\mathcal{A}$. Let $d : (\Theta \times \Theta, \mathcal{T} \otimes \mathcal{T}) \to (\mathbb{R}_+, \mathcal{B})$ be a non-negative measurable function where $\mathcal{B}$ is the Borel σ-algebra. Define

$$R = \inf_{\hat{\theta}} \max_{i\in\{1,2\}} E_i[d(\hat{\theta}, \theta_i)],$$

where $\inf_{\hat{\theta}}$ denotes the infimum with respect to all the measurable mappings $\hat{\theta} : (\mathcal{X}, \mathcal{A}) \to (\Theta, \mathcal{T})$, E_i denotes the expectation with respect to P_i, and θ_1, θ_2 are two elements of Θ.

Lemma 4 *Suppose that:*

(i) $d(\cdot,\cdot)$ *satisfies the triangle inequality,*

(ii) $\theta_1, \theta_2 \in \Theta$ *are such that* $d(\theta_1, \theta_2) \ge 2\psi$, *for some* $\psi > 0$,

(iii) $P_2 \ll P_1$ *and there exist constants* $\tau > 0$ *and* $0 < \gamma < 1$ *such that*

$$P_1\left[\frac{dP_2}{dP_1} \ge \tau\right] \ge 1 - \gamma.$$

Then

$$R \geq \psi(1-\gamma)\min\{\tau, 1\}. \tag{60}$$

Furthermore, if instead of (iii) *we suppose that*

(iv) $\chi^2(P_1, P_2) \leq \gamma^2$, *where* $0 < \gamma < 1$ *and*

$$\chi^2(P_1, P_2) = \int \left(\frac{dP_2}{dP_1} - 1\right)^2 dP_1,$$

then

$$R \geq \psi(1-\gamma)(1-\sqrt{\gamma}). \tag{61}$$

Proof. We first show (60). We have

$$\begin{aligned}
R &\geq \frac{1}{2}\inf_{\hat\theta}\Big(E_1[d(\hat\theta, \theta_1)] + E_2[d(\hat\theta, \theta_2)]\Big) \\
&\geq \frac{1}{2}\inf_{\hat\theta}\left(E_1[d(\hat\theta, \theta_1)] + \tau E_1\left[I\left(\frac{dP_2}{dP_1} \geq \tau\right) d(\hat\theta, \theta_2)\right]\right) \\
&\geq \frac{\min\{\tau, 1\}}{2}\inf_{\hat\theta} E_1\left[I\left(\frac{dP_2}{dP_1} \geq \tau\right)[d(\hat\theta, \theta_1) + d(\hat\theta, \theta_2)]\right].
\end{aligned}$$

Using here the triangle inequality and $(ii) - (iii)$, we find

$$R \geq \psi \min\{\tau, 1\} P_1\left[\frac{dP_2}{dP_1} \geq \tau\right] \geq \psi(1-\gamma)\min\{\tau, 1\}.$$

To show (61) it is sufficent to note that, in view of Chebyshev's inequality

$$P_1\left[\frac{dP_2}{dP_1} \geq 1 - \sqrt{\gamma}\right] = 1 - P_1\left[\frac{dP_2}{dP_1} - 1 < -\sqrt{\gamma}\right] \geq 1 - \frac{1}{\gamma}\int\left(\frac{dP_2}{dP_1} - 1\right)^2 dP_1 \geq 1 - \gamma,$$

and thus (iv) implies (iii) with $\tau = 1 - \sqrt{\gamma}$. □

Lemma 5 *For* $0 < \alpha, r, L < \infty$,

$$\sup_{f \in \mathcal{A}_{\alpha,r}(L)} \sup_{x \in \mathbb{R}} |f(x)| \leq L + \pi^{-1} C(r, \alpha),$$

where $C(r, \alpha) = \int_0^\infty \exp(-2\alpha u^r) du$.

Proof. Let $\Phi = \Phi^f$ be the characteristic function of f. Clearly,

$$|f(x)| \leq \frac{1}{2\pi}\int |\Phi(u)| du, \quad \forall x \in \mathbb{R}. \tag{62}$$

By Markov's inequality

$$\int |\Phi(u)|\, I\Big(|\Phi(u)| \exp(2\alpha|u|^r) > 1\Big) du \leq \int \exp(2\alpha|u|^r)\, |\Phi(u)|^2 du \leq 2\pi L.$$

Also,

$$\int |\Phi(u)|\ I\Big(|\Phi(u)| \exp\left(2\alpha|u|^r\right) \le 1\Big) du \quad \le \quad 2\int_0^\infty \exp\left(-2\alpha u^r\right) du = 2C(r,\alpha).$$

Combining the last two inequalities with (62) proves the Lemma. □

Lemma 6 *For any positive α, β, r, s and for any $A \in \mathbb{R}$ and $B \in \mathbb{R}$, we have*

$$\int_v^\infty u^A \exp\left(-\alpha u^r\right) du = \frac{1}{\alpha r} v^{A+1-r} \exp(-\alpha v^r)(1+o(1)), v \to \infty, \tag{63}$$

and

$$\int_0^v u^B \exp\left(\beta u^s\right) du = \frac{1}{\beta s} v^{B+1-s} \exp(\beta v^s)(1+o(1)), v \to \infty. \tag{64}$$

Proof. To prove (63), note first that, with $x = u^r$,

$$\int_v^\infty u^A e^{-\alpha u^r} du = \frac{1}{r}\int_{v^r}^\infty x^{(A+1)/r-1} e^{-\alpha x} dx \stackrel{\text{def}}{=} I.$$

Taking the integral by parts, we get

$$\begin{aligned} I &= -\frac{1}{\alpha r}\left[x^{(A+1)/r-1} e^{-\alpha x}\Big|_{v^r}^\infty - \left(\frac{A+1}{r} - 1\right)\int_{v^r}^\infty x^{(A+1)/r-2} e^{-\alpha x} dx\right] \\ &= \frac{1}{\alpha r} v^{A+1-r} e^{-\alpha v^r} + \frac{1}{\alpha r}\left(\frac{A+1}{r} - 1\right)\int_{v^r}^\infty x^{(A+1)/r-2} e^{-\alpha x} dx. \end{aligned}$$

Here,

$$\int_{v^r}^\infty x^{(A+1)/r-2} e^{-\alpha x} dx \le v^{-r}\int_{v^r}^\infty x^{(A+1)/r-1} e^{-\alpha x} dx = O(v^{-r} I) = o(I), v \to \infty.$$

Hence, as $v \to \infty$, we have

$$I(1+o(1)) = \frac{1}{\alpha r} v^{A+1-r} e^{-\alpha v^r},$$

which proves (63).

Let us prove (64). We have, with $x = u^r$,

$$\begin{aligned} \int_0^v u^B e^{\beta u^s} du &= \frac{1}{s}\int_0^{v^s} x^{(B+1)/s-1} e^{\beta x} dx \\ &= O\left(v^{(B+1)/2-s/2} e^{\beta v^{s/2}}\right) + I' = o(v^{B+1-s} e^{\beta v^s}) + I', \end{aligned}$$

as $v \to \infty$,where

$$I' = \frac{1}{s}\int_{v^{s/2}}^{v^s} x^{(B+1)/s-1} e^{\beta x} dx.$$

Taking the last integral by parts, we get

$$\begin{aligned} I' &= \frac{1}{\beta s}\left[x^{(B+1)/s-1}e^{\beta x}|_{v^{s/2}}^{v^s} - \frac{B+1-s}{s}\int_{v^{s/2}}^{v^s} x^{(B+1)/s-2}e^{\beta x}dx\right] \\ &= \frac{1}{\beta s}v^{B+1-s}e^{\beta v^s}(1+o(1)) + O(1)\int_{v^{s/2}}^{v^s} x^{(B+1)/s-2}e^{\beta x}dx, \end{aligned}$$

as $v \to \infty$. Here

$$\int_{v^{s/2}}^{v^s} x^{(B+1)/s-2}e^{\beta x}dx \le v^{-s/2}\int_{v^{s/2}}^{v^s} x^{(B+1)/s-1}e^{\beta x}dx = O(v^{-s/2}I') = o(I'),$$

as $v \to \infty$ and (64) follows. □

Lemma 7 *Let p be the density of stable symmetric distribution with characteristic function $\exp(-|t|^r)$, $1 < r < 2$. Then p is continuous, $p(x) > 0$ for all $x \in \mathbb{R}$ and there exist $c_1 > 0$, $c_2 > 0$ such that*

$$p(x) \ge c_1|x|^{-r-1},$$

for $|x| \ge c_2$.

Proof. From Zolotarev (1986), Th. 2.2.3., formula (2.2.18), we get

$$p(x) = \frac{r|x|^{1/(r-1)}}{2|1-r|}\int_0^1 u(\varphi)\exp(-|x|^{r/(r-1)}u(\varphi))d\varphi, \quad x \ne 0, \tag{65}$$

where

$$u(\varphi) = \left(\frac{\sin(\pi r\alpha/2)}{\cos(\pi\varphi/2)}\right)^{r/(1-r)}\frac{\cos(\pi(r-1)\varphi/2)}{\cos(\pi\varphi/2)}.$$

Clearly, for $\varphi \in [1/2, 1]$ we have

$$\begin{aligned} &1 \ge \cos(\pi(r-1)\varphi/2) \ge \cos(\pi(r-1)/4) > 0 \\ &c_3 \ge \sin(\pi r\varphi/2) \ge c_4 > 0, \end{aligned}$$

where $c_3 > 0$ and $c_4 > 0$ are constants. Thus,

$$c_6\left(\cos(\pi\varphi/2)\right)^{1/(r-1)} \le u(\varphi) \le c_5\left(\cos(\pi\varphi/2)\right)^{1/(r-1)},$$

$\varphi \in [1/2, 1]$, $c_5 > 0$, $c_6 > 0$ are constants. Now, if $\varphi \in [1/2, 1]$

$$c_7(1-\varphi) \le \cos(\pi\varphi/2) \le c_8(1-\varphi)$$

for some $c_7 > 0$, $c_8 > 0$. Finally,

$$c_{10}(1-\varphi)^{1/(r-1)} \le u(\varphi) \le c_9(1-\varphi)^{1/(r-1)}, \forall\varphi \in [1/2, 1].$$

Using (65) and the fact that $u(\varphi) \geq 0$ for $\varphi \in [0,1]$, we get

$$\begin{aligned} p(x) &\geq c|x|^{1/(r-1)} \int_{1/2}^{1} (1-\varphi)^{1/(r-1)} \exp\left(-|x|^{r/(r-1)} c_9 (1-\varphi)^{1/(r-1)}\right) d\varphi \\ &= c|x|^{1/(r-1)} \int_{0}^{1/2} \varphi^{1/(r-1)} \exp\left(-c_9 (|x|^r \varphi)^{1/(r-1)}\right) d\varphi. \end{aligned}$$

Here and further on $c > 0$ are constants, probably different on different occasions.

By change of variables, $u = (|x|^r \varphi)^{1/(r-1)}$, we get

$$\begin{aligned} p(x) &\geq c|x|^{1/(r-1)} \int_{0}^{(|x|^r/2)^{1/(r-1)}} \frac{u}{|x|^{r/(r-1)}} \exp(-c_9 u) \frac{u^{r-1}}{|x|^r} du \\ &= c|x|^{-1-r} \int_{0}^{(|x|^r/2)^{1/(r-1)}} u^{r-1} \exp(-c_9 u) du \\ &\geq c|x|^{-1-r} \int_{0}^{(c_2^r/2)^{1/(r-1)}} u^{r-1} \exp(-c_9 u) du \geq c_1 |x|^{-1-r}, \end{aligned}$$

for $|x| \geq c_2 > 0$. □

Lemma 8 *Let* $0 < r < s < \infty$ *and let* $h_* = h_*(n)$ *be defined by* (13), *i.e.*

$$\frac{2\alpha}{h_*^r} + \frac{2\beta}{h_*^s} = \log n - (\log\log n)^2.$$

Let h_n *satisfy*

$$b \log h_n + \frac{2\alpha}{h_n^r} + \frac{2\beta}{h_n^s} = \log n + C(1 + o(1)), \quad n \to \infty,$$

for some $b \in \mathbb{R}$ *and* $C \in \mathbb{R}$. *Then, as* $n \to \infty$, *we have*

$$h_*(n) = (\log n/(2\beta))^{-1/s}(1 + o(1)), \tag{66}$$

$$h_n^a \exp\left(-\frac{2\alpha}{h_n^r}\right) = h_*^a \exp\left(-\frac{2\alpha}{h_*^r}\right)(1 + o(1)), \tag{67}$$

$$\frac{h_*^a}{n} \exp\left(\frac{2\beta}{h_*^s}\right) = o\left(\exp\left(-\frac{2\alpha}{h_*^r}\right)\right), \tag{68}$$

for any $a \in \mathbb{R}$, *and*

$$h_*^{s-2\gamma-1} \exp\left(\frac{2\beta}{h_*^s}\right) \leq h_n^{s-2\gamma-1} \exp\left(\frac{2\beta}{h_n^s}\right), \tag{69}$$

for n *large enough.*

Proof. Define $x_* = h_*^{-s}$, $x_n = h_n^{-s}$, and write, for $t > 0$,

$$F(t) \stackrel{\text{def}}{=} 2\beta t + 2\alpha t^{r/s}, \quad F_1(t) \stackrel{\text{def}}{=} (-b/s)\log t + 2\beta t + 2\alpha t^{r/s}.$$

Then

$$\begin{aligned} F(x_*) &= \log n - (\log\log n)^2, &(70)\\ F_1(x_n) &= \log n + C(1+o(1)), &(71)\end{aligned}$$

for a constant $C \in \mathbb{R}$. We first prove that x_n satisfies

$$F(x_n) = 2\beta x_n + 2\alpha x_n^{r/s} = \log n + C_1 \log\log n(1+o(1)) + C_2(1+o(1)) \qquad (72)$$

for some constants $C_1, C_2 \in \mathbb{R}$. In fact,

$$F_1'(x_n) = \frac{1}{x_n}\left(-\frac{b}{s}\right) + 2\beta + \frac{2\alpha r}{s} x_n^{r/s-1} > 0,$$

for x_n large enough, thus $F_1(t)$ is strictly monotone increasing for large t, and a solution x_n of (71) exists for large n (and is unique). Next, clearly,

$$\frac{F_1(t)}{2\beta t} \to 1, \quad t \to \infty,$$

and therefore $\log n/(2\beta x_n) \to 1$, as $n \to \infty$. Similarly, $\log n/(2\beta x_*) \to 1$, as $n \to \infty$, which yields (66). Thus $(-b/s)\log x_n = (-b/s)\log\log n(1+o(1))$, as $n \to \infty$, and (72) follows. Next, let $F^{-1}(\cdot)$ be the inverse of $F(\cdot)$, then

$$\begin{aligned} (F^{-1}(x))' &= \frac{1}{F'(x)} = \frac{1}{2\beta + (2\alpha r/s)x^{r/s-1}},\\ (F^{-1}(x))'' &= \left(\frac{1}{2\beta + (2\alpha r/s)x^{r/s-1}}\right)' = \frac{-(2\alpha r/s)(r/s-1)x^{r/s-2}}{(2\beta + (2\alpha r/s)x^{r/s-1})^2}.\end{aligned}$$

We have

$$x_n = F^{-1}(\log n + a_n), \quad x_* = F^{-1}(\log n - b_n)$$

where $a_n = C_1 \log\log n(1+o(1)) + C_2(1+o(1)) = O(\log\log n)$ and $b_n = (\log\log n)^2$. Hence, for some $0 < \tau < 1$ and for n large enough,

$$\begin{aligned} x_n &= F^{-1}(\log n + a_n) = F^{-1}(\log n - b_n) + (F^{-1}(\log n - b_n))'(a_n + b_n)\\ &\quad + \frac{1}{2}\left(F^{-1}(\log n - b_n(1-\tau) + \tau a_n)\right)''(a_n + b_n)^2.\end{aligned}$$

Here

$$\begin{aligned} &(F^{-1}(\log n - b_n))' = \frac{1}{2\beta} + o(1), \quad n \to \infty,\\ &(F^{-1}(\log n - b_n(1-\tau) + \tau a_n))'' = O((\log n)^{r/s-2}), \quad n \to \infty.\end{aligned}$$

Thus,

$$\begin{aligned} x_* - x_n &= -\frac{1}{2\beta}(1+o(1))(a_n+b_n) + O\left(\frac{(a_n+b_n)^2}{(\log n)^{2-r/s}}\right) \\ &= -\frac{b_n}{2\beta}(1+o(1)) + O\left(\frac{(\log\log n)^4}{(\log n)^{2-r/s}}\right) \\ &= -\frac{b_n}{2\beta}(1+o(1)). \end{aligned} \tag{73}$$

Using this representation we obtain

$$\begin{aligned} &\exp\left(-\frac{2\alpha}{h_n^r} + \frac{2\alpha}{h_*^r}\right) = \exp(-2\alpha(x_n^{r/s} - x_*^{r/s})) \\ =\ & \exp\left(-2\alpha[x_* + b_n(2\beta)^{-1}(1+o(1))]^{r/s} - x_*^{r/s})\right) \\ =\ & \exp\left(-2\alpha x_*^{r/s}([1 + b_n(2\beta x_*)^{-1}(1+o(1))]^{r/s} - 1)\right) \\ =\ & \exp\left(O(b_n x_*^{r/s-1})\right) = 1 + o(1), \end{aligned}$$

since $b_n = (\log\log n)^2$, $x_* = (2\beta)^{-1}\log n\ (1+o(1))$ and $r < s$. This and the fact that $(h_n/h_*)^a = (x_*/x_n)^{sa} = 1 + o(1)$ imply (67). Next, (68) follows directly from the definition of h_* and from (66). To prove (69), note that, in view of (73),

$$\begin{aligned} &\frac{h_*^{s-2\gamma-1}}{h_n^{s-2\gamma-1}} \exp\left(\frac{2\beta}{h_*^s} - \frac{2\beta}{h_n^s}\right) = (1+o(1))\exp(2\beta(x_* - x_n)) \\ =\ & (1+o(1))\exp(-b_n[1+o(1)]) \le 1 \end{aligned}$$

for n large enough. □

Lemma 9 *Let $0 < r < s < \infty$ and let $h_+ = h_+(n)$ be the solution of (36). Then $h_+(n) = (\log n/(2\beta))^{-1/s}(1+o(1))$,*

$$h_+^a \exp\left(-\frac{2\alpha}{h_+^r}\right) = h_*^a \exp\left(-\frac{2\alpha}{h_*^r}\right)(1+o(1)), \ \text{as } n \to \infty \tag{74}$$

and

$$(\log n)^b n \exp\left(-\frac{2\alpha}{h_+^r} - \frac{2\beta}{h_+^s}\right) = o(1), \tag{75}$$

as $n \to \infty$, for any $a \in \mathbb{R}$, $b \in \mathbb{R}$.

Proof is analogous to that of Lemma 8. □

Acknowledgement. We would like to thank Fabienne Comte and Marie-Luce Taupin for a discussion of the results.

References

[1] Artiles, L.M. (2001) Adaptive minimax estimation in classes of smooth functions. PhD Thesis, University of Utrecht.

[2] Belitser, E., and Levit, B. (2001) Asymptotically local minimax estimation of infinitely smooth density with censored data. *Ann. Inst. Statist. Math.* **53** *289-306.*

[3] Brown, L.D., Low, M.G. and Zhao, L.H. (1997) Superefficiency in nonparametric function estimation. *Ann. Statist.* **25** *898-924.*

[4] Carroll, R. J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** *1184-1186.*

[5] Cavalier,L., Golubev, G.K., Lepski, O.V. and Tsybakov, A.B. (2003) Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory of Probability and its Applications* **48** *534-556.*

[6] Efromovich, S. (1997) Density estimation in the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92** *526-535.*

[7] Efromovich, S. and Koltchinskii, V. (2001) On inverse problems with unknown operators. *IEEE Trans. Inform. Theory* **47** *2876-2893.*

[8] Ermakov M.S. (1989) Minimax estimation of the solution of an ill-posed convolution type problem. *Problems of Information Transmission* **25** *191-200.*

[9] Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** *1257-1272.*

[10] Fan, J. (1991b) Global behavior of deconvolution kernel estimates. *Statist. Sinica* **1** *541-551.*

[11] Goldenshluger, A. (1999) On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** *907-926.*

[12] Golubev, G.K. and Khasminskii, R.Z. (2001) Statistical approach to Cauchy problem for Laplace equation. In: *State of the Art in Probability and Statistics, Festschrift for W.R. van Zwet* (M. de Gunst, C. Klaassen, A. van der Vaart, eds), IMS Lecture Notes Monograph Series **36** *419-433.*

[13] Ibragimov, I. A. and Hasminskii, R. Z. (1983) Estimation of distribution density. *J. Soviet Math.* **21** *40-57.*

[14] Masry, E. (1991) Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory* **37** *1105-1115.*

[15] Pensky, M. and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** *2033-2053.*

[16] Ritov, Y. (1987) On a deconvolution of normal distributions. Preprint. University of Berkeley.

[17] Ruymgaart, F.H. (1993) A unified approach to inversion problems in statistics. *Math. Methods of Statistics* **2** *130-146.*

[18] Stefanski, L. A. and Carroll, R. J. (1990) Deconvoluting kernel density estimators. *Statistics* **21** *169-184.*

[19] Tsybakov, A. B. (2000) On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris, Serie 1* **330** *835-840.*

[20] Tsybakov, A. B. (2004) *Introduction à l'estimation non-paramétrique.* Springer, Berlin-Heidelberg.

[21] Zhang, C.H. (1990) Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18** *806-831.*

[22] Zolotarev, V.M. (1986) *One-dimensional Stable Distributions*, Transl. AMS., Providence, Rhode Island.

[1]Laboratoire de Probabilités et Modèles
Aléatoires (UMR CNRS 7599),
Université Paris VI
4, pl.Jussieu, Boîte courrier 188,
75252 Paris, France
e-mail: tsybakov@ccr.jussieu.fr

[2]Modal'X, Université Paris X
200, avenue de la République
92001 Nanterre Cedex, France
e-mail: butucea@ccr.jussieu.fr