Deconvolution of supersmooth densities with smooth noise

Cristina BUTUCEA

Modal'X, Université Paris X, 200, avenue de la République, 92 001 Nanterre Cedex, France

cristina.butucea@u-paris10.fr

LPMA, Université Paris VI, 175, rue de Chevaleret, 75 013 Paris, France

Key words and phrases: Deconvolution density; minimax estimation; smooth noise; supersmooth functions.

Abstract: We estimate the common probability density of i.i.d. random variables that we observe with white noise. The unknown density belongs to some class of supersmooth functions. The errors in variables have an ordinary smooth distribution, that is their characteristic function decays polynomially asymptotically. In this problem, we evaluate the minimax pointwise and $\mathbb{L}_2$ risks.

Déconvolution de densités super régulières avec un bruit régulier

Résumé : Nous estimons la densité de probabilité commune de variables aléatoires i.i.d. observées avec un bruit blanc. La densité inconnue appartient à une classe de fonctions super-règulières. Les erreurs ont une distribution régulière, i.e. leur fonction caractéristique décroit de manière polynomiale asymptotiquement. Dans ce problème, nous évaluons les risques minimax ponctuel et en norme $\mathbb{L}_2$.

1. INTRODUCTION

Let us consider $X_1, \ldots, X_n$, i.i.d. random variables having common probability density f. Let always $\Phi^f(u) = \Phi(u) = \int \exp(-iux)f(x)dx$ denote the Fourier transform of the function f. Moreover, we assume that f belongs to the class of supersmooth densities of order r:

$$\mathcal{A}_{\alpha,r}(L) = \{f \text{ density} | \int |\Phi(u)|^2 \exp(2\alpha|u|^r)du \leq 2\pi L\}, \tag{1}$$

where $\alpha > 0$, $L > 0$ and $0 < r \leq 2$ are real numbers. Note that stable laws with autosimilarity index r belong to such a class for adequate values of α and L. The most encountered examples are in such classes, such as the Gaussian law ($r = 2$) or the Cauchy distribution ($r = 1$).

We want to estimate the unknown density from noisy observations

$$Y_i = X_i + \varepsilon_i, \ \ i = 1, \ldots, n,$$

with noise variables ε_i i.i.d. and independent of X_i, of known distribution having density f^ε. The methods described here are valid for a collection of problems with different noise distributions of the same type. We suppose here, that the noise is ordinary smooth, that is its characteristic function Φ^ε is asymptotically decreasing as a polynomial function:

$$\frac{b}{(1+|u|^2)^{s/2}} \leq |\Phi^\varepsilon(u)| \leq \frac{B}{(1+|u|^2)^{s/2}}, \text{ when } |u| \to \infty. \tag{2}$$

For technical reasons, we assume that the density of the noise f^ε is in $\mathbb{L}_1$ and in $\mathbb{L}_2$ and that its characteristic function Φ^ε is continuously differentiable.

We describe in this paper the minimax rates of convergence of the pointwise and $\mathbb{L}_2$ estimation risks, with an accent on the lower bounds.

Definition 1. Let $\hat{f}_n = \hat{f}_n(Y_1, \dots, Y_n)$ be an arbitrary estimator (based on observations $Y_1, \dots, Y_n$) of the unknown deconvolution density f. We define

1. the pointwise risk of the estimator $\hat{f}_n$ of f, at an arbitrary point x: $r(\hat{f}_n, f, x) = E_f[|\hat{f}_n(x) - f(x)|^2]$,

2. the $\mathbb{L}_2$ risk of the estimator $\hat{f}_n$ of f: $r(\hat{f}_n, f, \mathbb{L}_2) = E_f[\|\hat{f}_n - f\|_2^2]$,

where the expectation is taken with respect to the convolution model, with true underlying deconvolution density f.

Definition 2. A minimax rate of convergence φ_n for estimating a function f over the class $\mathcal{A}_{\alpha,r}(L)$ in (1) is such that:

1. there is an estimator $\hat{f}_n$ called optimal in the minimax sense attaining this rate, i.e.

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{A}_{\alpha,r}(L)} \varphi_n^{-2} r(\hat{f}_n, f, \cdot) \le C < \infty;$$

2. no other estimator can attain a better rate

$$\liminf_{n\to\infty} \inf_{f_n} \sup_{f\in\mathcal{A}_{\alpha,r}(L)} \varphi_n^{-2} r(f_n, f, \cdot) \ge c > 0,$$

for some constants $C, c > 0$ and where the infimum is taken over all possible estimators of f.

Minimax estimation of a density in a deconvolution model was intensively studied since Carroll & Hall (1988) who gave rates of convergence of the kernel deconvolution estimator. We can briefly cite for estimation over Hölder classes of functions with smooth and supersmooth noise, respectively, the papers by Stefanski & Carroll (1990), Fan (1991) and Masry (1991) (for mixing data). They used kernel estimators as well.

Efromovich (1997) gave efficient minimax estimators by projection (that is, he computed the exact asymptotic expression of the pointwise and $\mathbb{L}_2$ risks) over $\mathbb{L}_2$-periodic Sobolev classes, with supersmooth noise. Goldenshluger (1999) generalized the minimax rates to adaptive rates over more general $\mathbb{L}_p$ Sobolev classes in association to both smooth and supersmooth noise.

Deconvolution densities in analytic functions' class ($r = 1$) mixed with Cauchy type noise were estimated by Tsybakov (2000), in a discrete model. Surprisingly, a $\log n$ loss in the rate is inevitable when passing from minimax to adaptive estimation with $\mathbb{L}_2$ risk in this model.

The case where both the unknown density and the noise are supersmooth is separately solved in Butucea & Tsybakov (2002). Different behaviours are distinguished, associated to different convergence rates. We note that kernel estimators attain in some cases better rates then wavelet estimators in Pensky & Vidakovic (1999) and they are efficient in some cases.

In our context, upper bounds of $\mathbb{L}_2$ risk where already attained by wavelet estimators, over classes similar to $\mathcal{A}_{\alpha,r}(L)$ in Pensky & Vidakovic (1999). Therefore we describe briefly in this context how kernel estimators attain the same convergence rates, for both pointwise and $\mathbb{L}_2$ risks. Nevertheless, these rates are not known to be optimal unless lower bounds are given which is our main statement. We give here lower bounds for both pointwise and $\mathbb{L}_2$ risks.

There is some similarity between deconvolution with ordinary smooth noise and direct estimation of the generalized s-derivative of a density in $\mathcal{A}_{\alpha,r}(L)$.

Let us mention former results when estimating a density having infinitely many derivatives with direct observations $X_1, \dots, X_n$ available. Starting with Ibragimov & Hasminskii (1983) minimax rates were computed over classes of density functions with bounded analytic continuation in a strip around the real axis, corresponding to our class $\mathcal{A}_{\alpha,1}(L)$ ($r = 1$). The efficient estimation of such densities with pointwise risk and second order evaluations of the risks, together with efficient estimation of a derivative of order integer s were found by Golubev & Levit ($1996a, b$). Efficient estimation with $\mathbb{L}_2$ risk was given by Schipper (1996). Let us note that the constants we obtain in Corollary 1 below, for $s = 0$ and $r = 1$, are the same for pointwise risk in Golubev & Levit (1996*b*) and for the $\mathbb{L}_2$ risk in Schipper (1996). The case $s = 0$ in our problem corresponds indeed to no-noise, direct estimation and our estimator becomes the optimal sinc-kernel estimator.

Analytic densities were also considered by Belitser (1998) in the random-censorship model. For a good review of minimax results for analytic functions we refer to Ibragimov (2002).

Direct estimation over more general classes of infinitely differentiable functions similar to $\mathcal{A}_{\alpha,r}(L)$ was done in minimax and adaptive approach by Lepski & Levit (1998) in white noise model and by Artiles (2001) in density model and can be generalized to estimation of the s-derivative. The kernel estimators are different in deconvolution and direct estimation problems but they have almost the same bias which is much smaller than the variance. So, the same kernel estimator is optimal and the rate is given by the variance which is the same. Nevertheless, in deconvolution problem, the variance is mainly explained by the behaviour of the noise. Therefore, the techniques for proving the lower bounds are essentially different as far as the choice of our hypotheses is concerned.

We remark also that even though we can describe constants associated to the minimax upper bounds, we have not developped efficient lower bounds. Indeed, since the noise distribution gives the risk expression (rate and constant) the constants change for each noise distribution verifying the general assumption (2).

A slightly different assumption can be introduced for the supersmooth unknown density f, namely that

$$\int |\Phi(u)|^2 \left(u^2+1\right)^m \exp(2\alpha|u|^r)du \leq 2\pi L, \tag{3}$$

for some $m > 0$. This doesn't change anything for our results as it is stated in Remark 1.

If m is an integer and $f \in \mathcal{A}_{\alpha,r}(L)$ then the derivative $f^{(m)}$ verifies the inequality (3). Estimating the m-th derivative is a similar problem and the rate is given in Remark 2.

The next section presents the estimation procedure based on kernels attaining the minimax rates and evaluates the associated pointwise and $\mathbb{L}_2$ risks. The last section proves the lower bounds for these risks.

2. OPTIMAL ESTIMATION PROCEDURE

Let us consider the estimator:

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^n K_n\left(\frac{Y_i - x}{h}\right),$$

where $h \to 0$, $nh \to \infty$, when $n \to \infty$ and a kernel whose Fourier transform is

$$\Phi^{K_n}(u) = 1_{[-1,1]}(u)(\Phi^\varepsilon)^{-1}(u/h). \tag{4}$$

Note that for the no-noise problem $s = 0$, we get the sinc-kernel known to be optimal in direct estimation of supersmooth functions.

Theorem 1. *The above kernel estimator with bandwidth $h = (\log n/(2\alpha))^{-1/r}$ is optimal in the sense of Definition 2 for the rate*

$$\varphi_n^2 = \frac{1}{n}\left(\frac{\log n}{2\alpha}\right)^{(2s+1)/r},$$

with respect to both pointwise and $\mathbb{L}_2$ risks. More precisely, for $M^Y > 0$ defined in Lemma 1:

$$\sup_{f\in\mathcal{A}_{\alpha,r}(L)} \varphi_n^{-2} E_f[|\hat{f}_n(x) - f(x)|^2] \leq \frac{M^Y(x) + o(1)}{\pi b^2(2s+1)},$$

$$\sup_{f\in\mathcal{A}_{\alpha,r}(L)} \varphi_n^{-2} E_f[\|\hat{f}_n - f\|_2^2] \leq \frac{1 + o(1)}{\pi b^2(2s+1)}.$$

PROOF Let us start with the pointwise risk. We write the usual decomposition

$$E_f[|\hat{f}_n(x)-f(x)|^2] \le |E_f\hat{f}_n(x)-f(x)|^2 + E_f[|\hat{f}_n(x)-E_f\hat{f}_n(x)|^2].$$

Then, on one hand, the bias can be written in terms of characteristic functions, using the inverse Fourier transform: $f(x) = 1/(2\pi)\int \exp(iux)\Phi(u)du$. We obtain:

$$\begin{aligned}
|E_f\hat{f}_n(x)-f(x)|^2 &= \left|E_f\left(\frac{1}{h}K_n\left(\frac{Y_1-x}{h}\right)\right)-f(x)\right|^2 \\
&= \left|\frac{1}{h}K_n\left(\frac{\cdot}{h}\right)*f^Y(x)-f(x)\right|^2 \\
&\le \frac{1}{(2\pi)^2}\left(\int\left|\Phi^{K_n}(uh)\Phi^Y(u)-\Phi^X(u)\right|du\right)^2.
\end{aligned}$$

In the convolution model, $\Phi^Y=\Phi\cdot\Phi^\varepsilon$ and we use Cauchy- Schwarz inequality to get

$$\begin{aligned}
|E_f\hat{f}_n(x)-f(x)|^2 &\le \frac{1}{(2\pi)^2}\left(\int|1_{[-1,1]}(uh)-1|\cdot|\Phi^X(u)|du\right)^2 \\
&\le \frac{1}{(2\pi)^2}\left(\int_{|u|>1/h}|\Phi^X(u)|\exp(\alpha|u|^r)\exp(-\alpha|u|^r)du\right)^2 \\
&\le \frac{L}{2\pi}\int_{|u|>1/h}\exp(-2\alpha|u|^r)du \le \frac{L}{2\pi\alpha r}h^{r-1}\exp\left(-\frac{2\alpha}{h^r}\right).
\end{aligned}$$

On the other hand, for the variance of our estimator, use Lemma 2 below:

$$\begin{aligned}
Var_f\hat{f}_n(x) &= E_f[|\hat{f}_n(x)-E_f\hat{f}_n(x)|^2] \le \frac{1}{n}Var_f\left(\frac{1}{h}K_n\left(\frac{Y_1-x}{h}\right)\right) \\
&\le \frac{1}{nh}\left|\frac{1}{h}K_n^2\left(\frac{\cdot}{h}\right)*f^Y(x)\right| \le \frac{M^Y}{\pi b^2(2s+1)}\frac{1+o(1)}{nh^{2s+1}}.
\end{aligned}$$

By putting the bias and the variance together, minimising the mean square error of $\hat{f}_n(x)$ comes down to choosing $h=(\log n/(2\alpha))^{-1/r}$ which makes the bias infinitely smaller than the variance. Then

$$r(\hat{f}_n,f,x) \le \frac{M^Y}{\pi b^2(2s+1)}\varphi_n^2(1+o(1)).$$

Very similarly, for the $\mathbb{L}_2$ risk we have the same decomposition. By Plancherel formula we get for the $\mathbb{L}_2$ bias:

$$\begin{aligned}
\|E_f\hat{f}_n-f\|_2^2 &= \frac{1}{2\pi}\int|\Phi^{K_n}(uh)\Phi^Y(u)-\Phi^X(u)|^2du \\
&\le \frac{1}{2\pi}\int_{|u|\ge 1/h}|\Phi^X(u)|^2\exp(2\alpha|u|^r)\exp(-2\alpha/h^r)du \le L\exp\left(-\frac{2\alpha}{h^r}\right).
\end{aligned}$$

For the $\mathbb{L}_2$ variance

$$E_f[\|\hat{f}_n-E_f\hat{f}_n\|^2] \le \frac{1}{nh}E_f\left(\frac{1}{h}\int K_n^2\left(\frac{Y_1-x}{h}\right)dx\right) \le \frac{\|K_n\|_2^2}{nh} \le \frac{1+o(1)}{\pi b^2(2s+1)nh^{2s+1}}.$$

Thus, the same bandwidth is optimal, giving:

$$r(\hat{f}_n,f,\mathbb{L}_2) \le \frac{1}{\pi b^2(2s+1)}\varphi_n^2(1+o(1)).$$

□

REMARK 1. When estimating a density function verifing (3), only the bias changes, respectively:

$$|E_f\hat{f}_n(x) - f(x)|^2 \leq \frac{L}{2\pi}\int_{|u|>1/h}(u^2+1)^{-m}\exp(-2\alpha|u|^r)du \leq \frac{L}{2\pi\alpha r}h^{2m+r-1}\exp\left(-\frac{2\alpha}{h^r}\right)$$

and

$$\|E_f\hat{f}_n - f\|_2^2 \leq Lh^{2m}\exp\left(-\frac{2\alpha}{h^r}\right).$$

Then, the same bandwidth is optimal in this problem as well and we get the same convergence rate.

REMARK 2. As we already noted, if $f \in \mathcal{A}_{\alpha,r}(L)$ then $f^{(m)}$ checks the inequality (3), for an integer $m > 0$. The mean square error (MSE) and the mean integrated square error (MISE) for evaluating $f^{(m)}$ have the same bias as in Remark 1, but a dominating variance slightly larger $Var_f(\hat{f}_n^{(m)}(x)) \leq O(1)/(nh^{2s+2m+1})$. This gives a normalizing rate of order $(\log n)^{(2s+2m+1)/r}/n$.

COROLLARY 1. *If the noise distribution is exactly known to have $\Phi^\varepsilon(u) = (1+u^2)^{-s/2}$, then the above estimator with the same h and for the same φ_n^2 is such that*

$$\begin{aligned}
\limsup_{n\to\infty}\sup_{f\in\mathcal{A}_{\alpha,r}(L)}\varphi_n^{-2}E_f[|\hat{f}_n(x)-f(x)|^2] &\leq \frac{M^Y}{\pi(2s+1)},\\
\limsup_{n\to\infty}\sup_{f\in\mathcal{A}_{\alpha,r}(L)}\varphi_n^{-2}E_f[\|\hat{f}_n - f\|_2^2] &\leq \frac{1}{\pi(2s+1)}.
\end{aligned}$$

This is immediate by the previous theorem and precise computation of $\|K_n\|_2^2$.

LEMMA 1. *Probability density functions in the class $\mathcal{A}_{\alpha,r}(L)$ given in* (1) *are uniformly bounded, i.e.*

$$\sup_{f\in\mathcal{A}_{\alpha,r}(L)}\|f\|_\infty \leq M,$$

for some $M > 0$ depending only on α, r and L. Moreover, if f^ε is the density of a smooth noise in (2), *then the resulting convolution densities $f^Y = f * f^\varepsilon$ are uniformly bounded, i.e.*

$$\sup_{f\in\mathcal{A}_{\alpha,r}(L)}\|f\|_\infty \leq M^Y,$$

for some $M^Y > 0$ depending only on α, r, L, s and B.

PROOF We use inverse Fourier transform and Cauchy-Schwarz inequality:

$$|f(x)|^2 = \frac{1}{(2\pi)^2}\left|\int e^{-iux}\Phi(u)du\right|^2 \leq \frac{1}{(2\pi)^2}\int|\Phi(u)|^2e^{2\alpha|u|^r}du\int e^{-2\alpha|u|^r}du \leq \frac{L}{2\pi}\int e^{-2\alpha|u|^r}du$$

and this bound doesn't depend on f in $\mathcal{A}_{\alpha,r}(L)$. Similarly,

$$\left\|f^Y\right\|_\infty^2 \leq \frac{1}{(2\pi)^2}\left(\int|\Phi(u)\Phi^\varepsilon(u)|\,du\right)^2 \leq \frac{L}{2\pi}\int\frac{B^2}{(1+|w|^2)^s}\exp\left(-2\alpha\,|w|^r\right)dw,$$

which finishes the proof. □

LEMMA 2. *The kernel K_n defined in* (4) *is such that*

1. $\|K_n\|_2^2 \leq \dfrac{1+o(1)}{\pi b^2(2s+1)h^{2s}}$;
2. $\left|\dfrac{1}{h}K_n^2\left(\dfrac{\cdot}{h}\right) * f^Y(x)\right| \leq \|K_n\|_2^2(f^Y(x)+o(1))$.

PROOF 1. By Plancherel formula:

$$\begin{aligned}
\|K_n\|_2^2 &= \frac{1}{2\pi}\int |\Phi^{K_n}(u)|^2 du = \frac{1}{\pi}\int_0^1 \frac{du}{|\Phi^\varepsilon(u/h)|^2} \le \frac{1}{\pi b^2}\int_0^1 \left(1+\left(\frac{u}{h}\right)^2\right)^s du \\
&\le \frac{h}{\pi b^2}\left(\int_0^M (1+v^2)^s dv + \int_M^{1/h}(1+v^2)^s dv\right),
\end{aligned}$$

for some constant $M>0$ large enough. The first integral in the right-hand side term can be denoted by $C(M)$. Let us evaluate

$$I_s = \int_M^{1/h}(1+v^2)^s dv = \left[v(1+v^2)^s\right]_M^{1/h} - 2s(I_s - I_{s-1}).$$

Reiterated on s, this gives

$$I_s = \frac{1+o(1)}{(2s+1)h^{2s+1}} \text{ and } \|K_n\|_2^2 \le \frac{1+o(1)}{\pi b^2(2s+1)h^{2s}}.$$

2. Let us study first the asymptotic behaviour of K_n. We assumed that Φ^ε is a continuously differentiable function, then

$$K_n(x) = \frac{1}{2\pi}\int_{-1}^1 e^{ixu}\frac{du}{\Phi^\varepsilon(u/h)} = \left[\frac{e^{ixu}}{2\pi i x\Phi^\varepsilon(u/h)}\right]_{-1}^1 - \frac{1}{2\pi i x}\int_{-1}^1 e^{ixu}\frac{\Phi^{\varepsilon\prime}(u/h)}{h(\Phi^\varepsilon(u/h))^2}du,$$

meaning that $|K_n(x)| = O(|x|^{-1})$, as $|x|\to\infty$.

We integrate first over the interval $|hu|\le\epsilon$, for an ϵ which tends to 0 such that $\epsilon/h\to\infty$ when $n\to\infty$ (take e.g. $\epsilon = (\log\log n)^{-1}$). By continuity of our functions f^Y we get for a small $\delta>0$:

$$\left|\int_{|hu|\le\epsilon} K_n^2(u)\left[f^Y(x) - f^Y(x-hu)\right]du\right| \le \delta\|K_n\|_2^2.$$

By Lemma 1, densities f^Y are uniformly bounded, then we can deduce that

$$\left|\int_{|hu|>\epsilon} K_n^2(u)\left[f^Y(x) - f^Y(x-hu)\right]du\right| \quad\le\quad 2M^Y\int_{|hu|>\epsilon} K_n^2(u)\,du \le \frac{O(1)}{h^{2s}}\int_{|u|>\epsilon/h}\frac{du}{u^2}$$

and $\int_{|u|>\epsilon/h} 1/u^2du = 2h/\epsilon = o(1)$, by our choice. This finishes the proof as $\|K_n\|_2^2 = O\left(1/h^{2s}\right)$. □

3. LOWER BOUNDS

In order to prove that the previous rates are optimal, we show that no other estimator can achieve better rates. Because of the pointwise risk, one more point must be considered. Indeed, the pointwise minimax risk depends on the unknown $f(x)$. We must avoid having a different model at each n such that $f*f^\varepsilon(x)\to 0$ when $n\to\infty$. For this reason, we restrain our class of densities to

$$\mathcal{A}^\delta_{\alpha,r}(L) = \{f\in\mathcal{A}_{\alpha,r}(L) : f*f^\varepsilon(x)\ge\delta>0\},$$

for some fixed $\delta>0$ arbitrary small.

THEOREM 2. *The rate $\varphi_n^2 = (\log n/(2\alpha))^{(2s+1)/r}/n$ is minimax in the sense of Definition 2, i.e.:*

$$\liminf_{n\to\infty}\inf_{f_n}\sup_{f\in\mathcal{A}^\delta_{\alpha,r}(L)} \varphi_n^{-2} r(f_n,f,\cdot)\ge c,$$

where $r(f_n,f,\cdot)$ denotes successively the pointwise and the $\mathbb{L}_2$ risks of an arbitrary estimator f_n and $c>0$ depends only on δ,α,r and s.

PROOF Since we do not search exact lower bounds, for the $\mathbb{L}_2$ risk, we reduce the integration domain to a compact set $[a, b]$. We can check that $(f_n - f)^2$ is a uniformly bounded function, by a constant depending on n, α, r, s and L. We can write the $\mathbb{L}_2$ norm as as a limit of a Riemann sum, then:

$$E_f \|f_n - f\|_2^2 = \lim_{k\to\infty} \frac{b-a}{k} \sum_{i=1}^{k} E_f \left[(f_n(x_k) - f(x_k))^2 \right],$$

with $x_k = a + (b-a)i/k$. It is then enough to prove the lower bounds on the pointwise risk $r(f_n, f, x) \geq c\varphi_n^2$ for arbitrary x in order to deduce a lower bound on the $\mathbb{L}_2$ risk.

In order to do this, we apply the Van Trees inequality, following the lines of proof in Golubev & Levit (1996*b*). Let us introduce a parametric subfamily of hypothesis in our class of functions. We choose a symmetric centered stable law f_0 such that, for $R = \max\{r, 1\}$, $f_0 > 0$ belongs to the class $\mathcal{A}_{2^{R-1}\alpha,r}(L/2)$. Remark that f_0 is slightly more regular for the proof of the lower bounds.

The main idea is to choose functions f_θ as far apart as possible from f_0 at the estimation point x, such that the resulting models $\{f_\theta * f^\varepsilon\}_\theta$ be as close as possible to $f_0 * f^\varepsilon$ in some distance. In Van Trees inequality, the distance between models is of the order of the χ^2 distance. Due to the particular construction we describe later on, this distance is upper bounded by n times the variance of a certain perturbation function K. It is convenient therefore to take again the sinc-function as K. The price we have to pay is more computation to evaluate $|f_\theta - f_0|$ which is related to K via Fourier transformation.

For $\theta \in [-\theta_n, \theta_n]$ define the functions

$$f_\theta(x) = f_0(x) + \theta H(x),$$

such that for a function K and $\bar{K}(x) = \int K(x-y) f_0^Y(y)dy$

$$f_\theta^Y(y) = f_0^Y(y)(1 + \theta(K(x-y) - \bar{K}(x))).$$

Functions H and K are defined via their Fourier transforms Φ^H and Φ^K, by

$$\begin{aligned}
\Phi^K(u) &= 1_{[-1,1]}(uh), \\
\Phi^H(u) &= \frac{1}{2\pi\Phi^\varepsilon(u)} \int e^{ixw} \Phi_0^Y(u+w)\Phi^K(w)dw - \frac{\Phi_0(u)}{2\pi} \int e^{ixw} \Phi_0^Y(w)\Phi^K(w)dw,
\end{aligned}$$

such that $h \to 0$, $\theta_n \to 0$ when $n \to \infty$ and

$$\frac{\theta_n^2}{h^{2(s-r+1)}} \exp\left(\frac{2^R\alpha}{h^r}\right) = O(1). \tag{5}$$

We assume without loss of generality that we estimate the deconvolution density at point $x = 0$. We note that K is a symmetric function and we can write

$$\Phi^H(u) = \frac{1}{2\pi\Phi^\varepsilon(u)} \int \Phi_0^Y(u+w)\Phi^K(w)dw - \frac{\Phi_0(u)}{2\pi} \int \Phi_0^Y(w)\Phi^K(w)dw, \tag{6}$$

with $\bar{K} = \int K(y) f_0^Y(y)dy = 1/(2\pi) \int \Phi_0^Y(w)\Phi^K(w)dw$.

Let λ_0 be a probability density on $[-1, 1]$, such that $\lambda_0(-1) = \lambda_0(1) = 0$ and having finite Fisher information $I_0 = \int_{-1}^{1} (\lambda_0'(u))^2/(\lambda_0(u))du$. We consider $\lambda_n(u) = \lambda_0(u/\theta_n)/\theta_n$ defined on $[-\theta_n, \theta_n]$ to be specified later and having the Fisher information $I_n = I_{\lambda_n} = I_0/\theta_n^2$.

Then, for the pointwise risk, we can write

$$\begin{aligned}
r(f_n, f, x) &= \inf_{\hat{f}_n} \sup_{f \in \mathcal{A}^\delta_{\alpha,r}(L)} E_f \left[\left| \hat{f}_n(0) - f(0) \right|^2 \right] \\
&\geq \inf_{\hat{f}_n} \sup_{|\theta| \leq \theta_n} E_\theta \left[\left| \hat{f}_n(0) - f_\theta(0) \right|^2 \right] \\
&\geq \inf_{\hat{f}_n} \int_{[-\theta_n, \theta_n]} E_\theta \left[\left| \hat{f}_n(0) - f_\theta(0) \right|^2 \right] \lambda(\theta)d\theta.
\end{aligned}$$

We apply at this point the van Trees inequality (see Gill & Levit 1995) for the Bayesian risk in the deconvolution model and get

$$r(f_n, f, x) \geq \left(\int_{[-\theta_n,\theta_n]} \frac{\partial f_\theta(0)}{\partial \theta}\lambda(\theta)d\theta\right)^2 \left(n\int_{[-\theta_n,\theta_n]} I(\theta)\lambda(\theta)d\theta + I_n\right)^{-1}, \tag{7}$$

where $I(\theta)$ is the Fisher information of $(f_\theta * f^\varepsilon)_{\theta\in[-\theta_n,\theta_n]} = (f_\theta^Y)_{\theta\in[-\theta_n,\theta_n]}$ in the deconvolution model. On one hand, the numerator in (7) becomes $\partial f_\theta(0)/\partial\theta = H(0)$ and it doesn't depend on θ. As for the denominator, we write the Fisher information as given by Lemma 3 below

$$I(\theta) = \int \frac{(\partial f_\theta^Y(y)/\partial\theta)^2}{f_\theta^Y(y)} dy \leq \int K^2(y) f_0^Y(y) dy (1 + o(1)).$$

We obtain from (5), (7) and the results in Lemma 3 below the following lower bound

$$r(f_n, f, x) \geq \frac{(H(0))^2}{nI(\theta) + I_0\theta_n^{-2}} \geq O(1)\frac{(f_0^Y(0))^2/h^{2s+2}}{nf_0^Y(0)/h + I_0 h^{-2(s-r+1)}\exp(2^R\alpha/h^r)}.$$

At this point we choose

$$h = \left(\frac{\log n}{2^R\alpha} - \frac{2s+1}{2^R\alpha r}\log\left(\frac{\log n}{2^R\alpha}\right)\right)^{-1/r},$$

which is equivalent to $(\log n)^{-1/r}$ and it is such that

$$I_0 h^{-2(s-r+1)}\exp(2^R\alpha/h^r) \leq I_0\left(\frac{\log n}{2^R\alpha}\right)^{(2s-2r+2)/r} n\left(\frac{\log n}{2^R\alpha}\right)^{-(2s+1)/r} = o\left(\frac{n}{h}\right).$$

Thus we get the needed lower bounds

$$r_n \geq O(1) f_0^Y(0)\frac{(\log n)^{(2s+1)/r}}{n}$$

if we add the fact that under our assumption: $f_0 * f^\varepsilon(0) \geq \delta > 0$. □

Lemma 3. *For θ_n such that (5) holds, the auxiliary functions H and K are such that*

1. $H(0) \geq O(1) f_0^Y(0)/h^{s+1}$ *and* $\int H(u)du = 0$;
2. $f_\theta^Y(y) = f_0^Y(y)(1 + o(1))$, *where* $o(1) \to 0$, *when* $n \to \infty$, *uniformly in* y *and* $|\theta| \leq \theta_n$;
3. *functions* f_θ *are density functions belonging to* $\mathcal{A}_{\alpha,r}(L)$;
4. $I(\theta) \leq \int K^2(y) f_0^Y(y)dy \leq f_0^Y(0)\|K\|_2^2(1 + o(1)) \leq f_0^Y(0)(1 + o(1))/(2\pi h)$.

Proof 1. By construction, in (6),

$$\begin{aligned} H(0) &= \frac{1}{2\pi}\int \Phi^H(u)du \geq \frac{1}{2\pi}\int\int 1_{[-1,1]}(h(u-w))\Phi_0^Y(w)/\Phi^\varepsilon(u)dwdu \\ &\geq \frac{1}{2\pi}\int \Phi_0^Y(w)\int_{w-1/h}^{w+1/h}(1+u^2)^{s/2}dudw \\ &\geq O(1)\int \Phi_0^Y(w)dw\int_{-1/h}^{1/h}(1+u^2)^{s/2}du \geq O(1)\frac{f_0^Y(0)}{h^{s+1}}. \end{aligned}$$

We have to check also that

$$\frac{1}{2\pi}\int\frac{\Phi_0(u)}{2\pi}\int \Phi_0^Y(w)\Phi^K(w)dwdu = f_0(0)\bar{K} = o\left(\frac{f_0^Y(0)}{h^{s+1}}\right).$$

Indeed,

$$\begin{aligned}
\bar{K} &= \int \Phi_0(w)\Phi^\varepsilon(w)\Phi^K(w)dw \\
&\leq \frac{B}{2\pi}\int_{|w|\leq 1/h}(1+w^2)^{-s/2}\Phi_0(w)dw \\
&\leq \frac{B\sqrt{L}}{2\pi}\left(\int_{|w|\leq 1/h}(1+w^2)^{-s}e^{-2\alpha|w|^r}dw\right)^{1/2} \\
&\leq O(1)h^{s+(r-1)/2}e^{-\alpha/h^r} = o(1),
\end{aligned}$$

which is an $o\left(1/h^{s+1}\right)$, too.

Moreover, $\int H(u)du = \Phi^H(0)/(2\pi) = 0$.

2. Indeed, by construction

$$|\theta K(y)| \leq \theta_n \|K\|_\infty \leq \theta_n \int \Phi^K(u)du = O\left(\frac{\theta_n}{h}\right) = o(1).$$

Remark that the densities f_0^Y in our class are uniformly bounded (see Lemma 1), that is $\|f_0^Y\|_\infty \leq O(1)$. Finally, by previous results of this Lemma

$$|\theta\bar{K}| \leq o(\theta_n) = o(1).$$

3. We need to check that f_θ is a positive function for n large enough, it is summable to 1 and it belongs to our class of functions. It is easy to check that it integrates to 1. Let us write now (see (6))

$$f_\theta(x) = f_0(x)(1-\theta\bar{K}) + \theta G(x),$$

where G has Fourier transform

$$\Phi^G(u) = \frac{1}{2\pi\Phi^\varepsilon(u/h)}\int_{|w|\leq 1}\Phi_0(u+w)\Phi^\varepsilon(u+w)dw.$$

We can say that Φ^G is smoother than Φ_0 because of the convolution with the identity function in the integral above. Roughly speaking, that means G is decreasingly faster than f_0 when $|x| \to \infty$ and thus f_θ is a positive function as soon as n is large enough.

At last, f_θ belongs to our class of function (use the generalized Minkowski inequality):

$$\begin{aligned}
& \theta^2 \int |\Phi^H(u)|^2 e^{2\alpha|u|^r}du \\
\leq\ & \theta_n^2 \int |\int \Phi_0^Y(u+w)\Phi^K(w)dw|^2(\Phi^\varepsilon(u))^{-2}e^{2\alpha|u|^r}du + \theta_n^2\bar{K}^2 2\pi L \\
\leq\ & \theta_n^2(\int(\int(\Phi_0^Y(u+w)/\Phi^\varepsilon(u))^2 e^{2\alpha|u|^r}du)^{1/2}\Phi^K(w)dw)^2 + o(1) \\
\leq\ & \theta_n^2(\int(\int \Phi_0^2(u)e^{2^R\alpha|u|^r}du)^{1/2}e^{2^{R-1}\alpha|w|^r}/\Phi^\varepsilon(w)\Phi^K(w)dw)^2 + o(1) \\
\leq\ & \theta_n^2 2\pi L(\int_{|w|\leq 1/h} B(1+w^2/2)^{s/2}e^{2^{R-1}\alpha|w|^r}dw)^2 + o(1) \\
\leq\ & O(1)\frac{\theta_n^2}{h^{2(s-r+1)}}\exp\left(\frac{2^R\alpha}{h^r}\right) + o(1) < \infty,
\end{aligned}$$

and this constant can be smaller than πL by our choice of θ_n.

We used the facts that $e^{2\alpha|u|^r} \leq e^{2^R|u+w|^r}e^{2^R|w|^r}$, where we recall that $R = \max\{r, 1\}$ and that

$$\frac{(1+|u|^2)^s}{(1+|u+w|^2)^s} \leq (1+|w|^2/2)^s.$$

4. The crucial point is to see that $\int K^2(y)(f_0^Y(y) - f_0^Y(0))dy = o(1/h)$. We split this integral into two parts, on a neighbourhood of 0 denoted $O_\epsilon(0)$, respectively on the remaining part, where $h/\epsilon = o(1)$. Then

$$\int_{O_\epsilon(0)} K^2(y)(f_0^Y(y) - f_0^Y(0))dy \leq o(1)\|K\|_2^2 = o(1/h),$$

by continuity of f_0^Y. On the other interval note again that the densities in our class are uniformly bounded (see Lemma 1), then we write

$$|\int_{O_\epsilon^C(0)} K^2(y)(f_0^Y(y) - f_0^Y(0))dy| \leq O(1)\int_{|y|\geq\epsilon} \frac{sin^2(y/h)}{\pi^2 y^2}dy \leq O(1/\epsilon)$$

and this is an $o(1/h)$ by construction. □

ACKNOWLEDGEMENTS

The author is very grateful to the anonymous referee and the associate editor whose very kind advice helped improving the presentation of this work.

REFERENCES

L. M. Artiles (2001). Adaptive minimax estimation in classes of smooth functions. *PhD Thesis, University of Utrecht*, ISBN: 90-3932-903-6.

E. Belitser (1998) Efficient estimation of analytic density under random censorship. *Bernoulli*, 4, 519–543.

C. Butucea & A. B. Tsybakov (2003). Fast asymptotics in density deconvolution. *Manuscript.*

R. J. Carroll & P. Hall (1988) Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, 83, 1184–1186.

S. Efromovich (1997) Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.*, 92, 526–535.

J. Fan (1991) On the optimal rate of convergence for nonparametric deconvolution problem. *Ann. Statist.*, 19, 1257–1272.

R. D. Gill & B. Y. Levit (1995). Application of the van Trees inequality: a Bayesian Cramer-Rao bound. *Bernoulli*, 1, 59–79.

A. Goldenshluger (1999) On pointwise adaptive nonparametric deconvolution. *Bernoulli*, 5, 907–926.

G. K. Golubev & B. Y. Levit (1996a) On the second order minimax estimation of distribution functions. *Mathem. Meth. Statist.*, 5, 1–31.

G. K. Golubev & B. Y. Levit (1996b) Asymptotically efficient estimation for analytic distributions. *Mathem. Meth. Statist.*, 5, 357–368.

I. A. Ibragimov & R. Z. Hasminskii (1983) Estimation of distribution density. *J. Soviet Math.*, 25, 40–57.

I. Ibragimov (2001) Estimation of analytic functions. *State of the art in probability and statistics (Leiden, 1999), IMS Lecture Notes Monogr. Ser., 36*, 359–383,

O. V. Lepski & B. Y. Levit (1998) Adaptive minimax estimation of infinitely differentiable functions. *Mathem. Meth. Statist.*, 7, 123–156.

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

M. Pensky & B. Vidakovic (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27, 2033–2053.

M. Schipper (1996) Optimal rates and constants in $\mathbb{L}_2$-minimax estimation of probability density functions *Mathem. Meth. Statist.*, 5, 253-274.

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

A. B. Tsybakov (2000) On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci. Paris*, 330, Série I, 835–840.