Sharp optimality for density deconvolution with dominating bias. I

Cristina BUTUCEA\textsuperscript{1,2} and Alexandre B. TSYBAKOV\textsuperscript{1}
\textsuperscript{1}Université Paris VI, \textsuperscript{2}Université Paris X

November 15, 2006

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 whose variance turns out to be asymptotically negligible w.r.t. its squared bias both under the pointwise and \( L_2 \)-risks.

We prove in Part II that this estimator is optimal in sharp asymptotical minimax sense on \( \mathcal{A} \) simultaneously under the pointwise and the \( L_2 \)-risks. For \( r < s/2 \) we construct a sharp adaptive estimator of \( f \).

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. I.

1 Introduction

Assume that one observes \( Y_1, \ldots, Y_n \) in the model

\[ Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, n, \]

where \( X_i \) are i.i.d. random variables with an unknown probability density \( f \) w.r.t. the Lebesgue measure on \( \mathbb{R} \), the random variables \( \varepsilon_i \) are i.i.d. with known probability density \( f^\varepsilon \) w.r.t. the Lebesgue measure on \( \mathbb{R} \), and \((\varepsilon_1, \ldots, \varepsilon_n)\) is independent of \((X_1, \ldots, X_n)\). The deconvolution problem that we consider here is to estimate \( f \) from observations \( Y_1, \ldots, Y_n \).
Denote by \( f^Y = f \ast f^\varepsilon \) the density of the variables \( Y_i \), where \( \ast \) is the convolution sign. Let \( \Phi^Y, \Phi^X \) and \( \Phi^\varepsilon \) be the characteristic functions of random variables \( Y_i, X_i \) and \( \varepsilon_i \), respectively. For an integrable function \( g : \mathbb{R} \to \mathbb{R} \), define the Fourier transform

\[
\Phi^g(u) = \int g(x) \exp(iux) du.
\]

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

\[
\mathcal{A}_{\alpha,r}(L) = \{ f : \mathbb{R} \to \mathbb{R}_+ }, \int f = 1 \text{ and } \int |\Phi^f(u)|^2 \exp(2\alpha|u|^r) du \leq 2\pi L, \]

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 \( \varepsilon \) satisfies the following assumption.

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

\[
b_{\text{min}}|u|\gamma \exp(-\beta |u|^s) \leq |\Phi^\varepsilon(u)| \leq b_{\text{max}}|u|\gamma' \exp(-\beta |u|^s) \tag{1}
\]

for \( |u| \geq u_0 \).

Many important probability densities belong to the class \( \mathcal{A}_{\alpha,r}(L) \) with some \( \alpha, r, L \) 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 \( f^\varepsilon \) that should be explicitly known. Therefore, without a meaningful loss, we will sometimes restrict our study to the case \( 0 < s \leq 2 \).

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

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

and the maximal \( L_2 \)-risk

\[
R_n(L_2, \hat{f}_n, \mathcal{A}_{\alpha,r}(L)) = \sup_{f \in \mathcal{A}_{\alpha,r}(L)} E_f \left[ \|\hat{f}_n - f\|_2^2 \right],
\]

2
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 \( L_2(\mathbb{R}) \)-norm. (In what follows we use the notation \( L_p(\mathbb{R}) \), in general, for the \( L_p \)-spaces of complex valued functions on \( \mathbb{R} \).)

The asymptotics of minimax risks differ significantly for the cases \( r < s \), \( r = s \) and \( r > s \). If \( r < s \) the "variance" part of the asymptotical minimax risks is asymptotically negligible w.r.t. the "bias" part, while for \( r > s \) the bias contribution 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 \( A_{\alpha,r}(L) \). 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 \( A_{\alpha,r}(L) \), 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) were the first to point out the effect of fast rates in density deconvolution, considering the classes of densities that are somewhat smaller than \( A_{\alpha,r}(L) \) (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 \( L_2 \)-risk. Our results imply that the rates achieved by their estimators are suboptimal on \( A_{\alpha,r}(L) \) and that the optimal rates can be attained by a simpler and more traditional kernel deconvolution method with suitably chosen parameters. We will show in Part II of this paper, see Butucea and Tsybakov (2007), 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 \( L_2 \)-risk and under the pointwise risk and that it is sharp adaptive to the parameters \( \alpha, r, L \) in some cases.

The most difficult part of our results is the construction of minimax
lower bounds. The technique that we develop in Part II, Butucea and Tsybakov (2007), 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 $f^\varepsilon$ are both periodic on $[0,2\pi]$, $f^\varepsilon$ 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 $f$:

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

(2)

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

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

(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 $\Phi_{K_n} \in L_2(\mathbb{R})$ (and thus $\Phi_{K_n} \in L_1(\mathbb{R})$). This holds under mild assumptions. For example, in view of the continuity property of characteristic functions, the assumption that $\Phi^\varepsilon(u) \neq 0$ for all $u \in \mathbb{R}$ is sufficient to have $\Phi_{K_n} \in L_2(\mathbb{R})$. Moreover, the condition $\Phi_{K_n} \in L_2(\mathbb{R})$ implies that the kernel $K_n$ is real-valued. In fact, under this
condition we have $\Phi_{K_n}(u) = \Phi(\epsilon(-u/h_n))V_n(u)$ for almost all $u \in \mathbb{R}$, where $V_n(u) = I(|u| \leq 1)/|\Phi(\epsilon(u/h_n))|^2$ is an even real-valued function belonging to $L_1(\mathbb{R})$ and $\Phi(\epsilon(-u/h_n))$ (the complex conjugate of $\Phi(\epsilon(u/h_n))$) is the Fourier transform of real-valued function $t \mapsto h_n f(\epsilon(-h_n t))$. 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 $L_2$ bias terms, i.e. on the quantities $|E f \hat{f}_n(x) - f(x)|^2$ and $\|E f \hat{f}_n - f\|^2$.

**Proposition 1** Let $f \in A_{\alpha,r}(L)$, $\alpha > 0, r > 0, L > 0$ and assume that $\Phi_{K_n} \in L_2(\mathbb{R})$ for any $h_n > 0$. Then the squared bias of $\hat{f}_n(x)$ is bounded as follows

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

as $h_n \to 0$, while the bias term of the $L_2$-risk satisfies

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

for every $h_n > 0$.

**Proof.** For the pointwise bias we have

$$|E f \hat{f}_n(x) - f(x)|^2 = \left|\left(\frac{1}{h_n} K_n \left(\frac{\cdot}{h_n}\right) * f^Y(\cdot)\right)(x) - f(x)\right|^2$$

$$= \left|\frac{1}{2\pi} \int [\Phi_{K_n}(uh_n)\Phi^Y(u) - \Phi^X(u)] \exp(-iux)du\right|^2$$

$$\leq \frac{1}{(2\pi)^2} \left(\int I(|uh_n| > 1)|\Phi^X(u)|^2 \right)^2.$$

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

$$\left|E f \hat{f}_n(x) - f(x)\right|^2$$

$$\leq \frac{1}{(2\pi)^2} \int_{|u| > 1/h_n} \exp(-2\alpha|u|^r)du \int_{|u| > 1/h_n} |\Phi^X(u)|^2 \exp(2\alpha|u|^r)du$$

$$\leq L \frac{1}{2\pi} \int_{|u| > 1/h_n} \exp(-2\alpha|u|^r)du.$$
which together with Lemma 2 yields the first inequality of the Proposition. To prove the second inequality, we apply the Plancherel formula and get
\[
\|Ef\hat{f} - f\|_2^2 = \left\| \frac{1}{h_n} E_f K_n \left( \frac{\cdot - Y_1}{h_n} \right) - f(\cdot) \right\|_2^2 \\
= \frac{1}{2\pi} \int |\Phi^*_n(uh_n)\Phi(u) - \Phi^*(u)|^2 du \\
= \frac{1}{2\pi} \int I(|uh_n| > 1)|\Phi^*(u)|^2 du \\
\leq \frac{\exp(-2\alpha/h^*_n)}{2\pi} \int_{|u|>1/h_n} |\Phi^*(u)|^2 \exp(2\alpha|u^*|) du.
\]
(5)

The next proposition gives upper bounds on the pointwise and the $L^2$ variance terms defined as

\[
\text{Var}_f \hat{f}_n(x) = E_f \left[ |\hat{f}_n(x) - E_f \hat{f}_n(x)|^2 \right] \quad \text{and} \quad \text{Var}_{f,2} \hat{f}_n = E_f \left[ \|\hat{f}_n - E_f \hat{f}_n\|_2^2 \right]
\]
respectively.

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

\[
\sup_{x \in \mathbb{R}} \text{Var}_f \hat{f}_n(x) = \sup_{x \in \mathbb{R}} E_f \left[ |\hat{f}_n(x) - E_f \hat{f}_n(x)|^2 \right] \\
\leq \min \left( f^*, \frac{4}{\beta \bar{s} h_n^s - 1} \right) \frac{h_n^{s+2\gamma-1}}{2\pi \beta \bar{s} b_{\min}^2 n} \exp \left( \frac{2\beta}{h_n^*} \right) (1 + o(1))
\]
(6)

as $h_n \to 0$, and, for an arbitrary density $f$, the variance term of the $L^2$-risk satisfies

\[
\text{Var}_{f,2} \hat{f}_n = E_f \left[ \|\hat{f}_n - E_f \hat{f}_n\|_2^2 \right] \leq \frac{h_n^{s+2\gamma-1}}{2\pi \beta \bar{s} b_{\min}^2 n} \exp \left( \frac{2\beta}{h_n^*} \right) (1 + o(1))
\]
(7)

as $h_n \to 0$.

**Proof.** For the pointwise variance we obtain two separate bounds and then take the minimum of them. To get the first bound, we write

\[
\text{Var}_f \hat{f}_n(x) = \frac{1}{n} E_f \left[ \left| \frac{1}{h_n} K_n \left( \frac{x - Y_1}{h_n} \right) - E_f \left[ \frac{1}{h_n} K_n \left( \frac{x - Y_1}{h_n} \right) \right] \right|^2 \right] \\
\leq \frac{1}{n h_n} \int \frac{1}{h_n} K_n^2 \left( \frac{x - y}{h_n} \right) f^*(y) dy \\
\leq \frac{f^*}{n h_n} \|K_n\|_2^2.
\]
(8)
where we used the fact that the convolution density \( f^Y = f \ast f^\varepsilon \) is uniformly bounded by \( f_\ast \). Applying the Plancherel formula and using (1) and (25) of Lemma 2 in the Appendix we get

\[
\|K_n\|^2_2 = \frac{h_n}{2\pi} \int_{|u| \leq 1/h_n} |\Phi^\varepsilon(u)|^{-2} du
\]

\[
\leq \frac{h_n}{2\pi b_{\min}^2} \int_{u_0 \leq |u| \leq 1/h_n} |u|^{-2\gamma} \exp(2\beta|u|^s) du + \frac{h_n}{2\pi} \int_{|u| \leq u_0} |\Phi^\varepsilon(u)|^{-2} du
\]

\[
\leq \frac{h_n}{\pi s^2 \min \beta^2} \int_0^{1/h_n} |u|^{-2\gamma} \exp(2\beta u^s) du + O(h_n)
\]

\[
= \frac{h_n^{s+2\gamma}}{2\pi b_{\min}^2} \exp \left( \frac{2\beta}{h_n^s} \right) (1 + o(1)), \quad h_n \to 0.
\]  

(9)

This and (8) imply the first bound in (6). For the second bound we still use the second line in (8) but then we apply the Plancherel formula in a different way:

\[
Var_f \hat{f}_n(x) \leq \frac{1}{2\pi n} \int \left| \Phi^{K_1,n}(u) \Phi^{Y,Y}(u) \right| du \leq \frac{1}{2\pi n} \int \left| \Phi^{K_1,n}(u) \right| du,
\]

where \( K_{1,n}(\cdot) = K_n(\cdot/h_n)/h_n \) and \( \Phi^Y \) is the complex conjugate of \( \Phi^Y \). Thus, using that \( \Phi^{K_{1,n}}(u) = \Phi^{K_n(h_n u)} \) and then acting similarly to (9) we get

\[
Var_f \hat{f}_n(x) \leq \frac{1}{2\pi n} \int \left| \Phi^{K_{1,n} \ast \Phi^{K_{1,n}}}(u) \right| du \leq \frac{1}{2\pi n} \left( \int \left| \Phi^{K_n}(h_n u) \right| du \right)^2
\]

\[
= \frac{1}{2\pi n} \left( \int_{|u| \leq 1/h_n} |\Phi^\varepsilon(u)|^{-1} du \right)^2
\]

\[
\leq \frac{2h_n^{2s+2\gamma-2}}{\pi \beta^2 s^2 b_{\min}^2} \exp \left( \frac{2\beta}{h_n^s} \right) (1 + o(1)),
\]

which yields the second bound in (6). Finally,

\[
Var_{f_2} \hat{f}_n \leq \frac{1}{nh_n} \int \int \frac{1}{h_n} K_n^2 \left( \frac{x-y}{h_n} \right) f^Y(y) dy dx
\]

\[
= \frac{1}{nh_n} \|K_n\|^2_2,
\]

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

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

\[
f_\ast = \sup_{f \in A_{\alpha,r}(L)} \sup_{x \in \mathbb{R}} |f(x)|.
\]
This value is finite and can be taken as in Lemma 1 of the Appendix.

Interestingly, inequality (6) shows that asymptotics of the pointwise variance are different for $0 < s < 1$ and $s > 1$, while this is not the case for the $L_2$ variance term given by (7). Inequality (6) can be compared to the recent result of van Es and Uh (2004). They studied asymptotic pointwise variance of the same deconvolution kernel estimator in the particular case of stable noise distributions with $1/3 < s \leq 2$ and also noticed that $s = 1$ marks a change of behaviour. These effects concerning variance terms will not be crucial in what follows since we will consider the bias dominated case.

Let us note that for practical implementation we can improve the kernel by smoothing its Fourier transform. One possibility is the following: let $K_n$ be the kernel defined via its Fourier transform

$$
\Phi_{K_n}(u) = \frac{\Psi(u)}{\Phi^x(u/h_n)}
$$

$$
\Psi(u) = I(|u| \leq 1 - \delta_n) + \frac{1 - |u|}{\delta_n} I(1 - \delta_n \leq |u| \leq 1),
$$

for some $\delta_n \to 0$ such that $\delta_n/h_n^r \to 0$ and $\delta_n/h_n^s \to 0$, as $n \to \infty$. Then $|K_n(x)| = O(1/|x|^2)$ as $|x| \to \infty$, $K_n$ is integrable for any $n$ and the resulting estimator is also integrable. Moreover, by further smoothing the function $\Psi$ we can obtain a Fourier transform $\Phi_{K_n}$ as smooth as $\Phi^x$ and the resulting kernel estimator will have as many finite absolute moments as the noise distribution.

3 Optimal bandwidths and upper bounds for the risks

Propositions 1 and 2 lead to upper bounds for pointwise and $L_2$ 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

\[ E_f \left[ \left( \hat{f}_n(x) - f(x) \right)^2 \right] = \left| E_f \hat{f}_n(x) - f(x) \right|^2 + Var_f \hat{f}_n(x) \]

\[ \leq \frac{L}{2\pi \alpha \gamma} h_n^{-1} \exp \left( \frac{-2\alpha}{h_n^r} \right) + \frac{f_s}{2\pi \beta s b^2_{\min}} \frac{h_n^{s+2\gamma-1}}{n} \exp \left( \frac{2\beta}{h_n^s} \right). \]

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

\[ \frac{L b^2_{\min}}{f_s} n \tilde{h}_n^{-2\gamma} (1 + o(1)) = \exp \left( \frac{2\alpha}{\tilde{h}_n^r} + \frac{2\beta}{\tilde{h}_n^s} \right), \]  

(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 \( h_n \) is a solution in \( h_n \) of the equation

\[ 2\gamma \log h + \frac{2\alpha}{h_n^r} + \frac{2\beta}{h_n^s} = \log n + C (1 + o(1)), \]  

(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

\[ \tilde{h}_n^{-r} \exp \left( \frac{-2\alpha}{\tilde{h}_n^r} \right) = C (1 + o(1)) \frac{\tilde{h}_n^{r+2\gamma-1}}{n} \exp \left( \frac{2\beta}{\tilde{h}_n^s} \right), \]

\[ = C (1 + o(1)) \tilde{h}_n^{-s} \frac{\tilde{h}_n^{s+2\gamma-1}}{n} \exp \left( \frac{2\beta}{\tilde{h}_n^s} \right), \]

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 \( L_2 \)-risk we get

\[ E_f \left[ \| \hat{f}_n - f \|_2^2 \right] = \| E_f \hat{f}_n - f \|_2^2 + Var_{f,2} \hat{f}_n \]

\[ \leq L \exp \left( \frac{-2\alpha}{h_n^r} \right) + \frac{1}{2\pi \beta s b^2_{\min}} \frac{h_n^{s+2\gamma-1}}{n} \exp \left( \frac{2\beta}{h_n^s} \right), \]

and the minimizer \( h_n = h_n(\|2\) of the last expression is a solution in \( h \) of the equation

\[ (r + 2\gamma - 1) \log h + \frac{2\alpha}{h_n^r} + \frac{2\beta}{h_n^s} = \log n + C (1 + o(1)). \]  

(12)
Now, this equation implies
\[
\exp \left( -\frac{2\alpha}{h_n^r(\mathbb{L}_2)} \right) = C(1 + o(1)) \frac{h_n^{r+2\gamma-1}(\mathbb{L}_2)}{n} \exp \left( \frac{2\beta}{h_n^s(\mathbb{L}_2)} \right) = C(1 + o(1)) h_n^{r-s}(\mathbb{L}_2) \frac{h_n^{s+2\gamma-1}(\mathbb{L}_2)}{n} \exp \left( \frac{2\beta}{h_n^s(\mathbb{L}_2)} \right),
\]
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
\[
\frac{2\beta}{h_*^s} + \frac{2\alpha}{h_*^r} = \log n - (\log \log n)^2,
\]
(in what follows we will assume w.l.o.g. that \( n \geq 3 \) to ensure that \( \log n > (\log \log n)^2 \)). Lemma 4 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 \( h_n \) 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 \( h_n \) and \( h_n(\mathbb{L}_2) \) respectively.

Note that, in fact, \( h_* \) is better than both bandwidths \( h_n \) and \( h_n(\mathbb{L}_2) \) in the variance terms, but these terms are asymptotically negligible w.r.t. the bias ones (cf. Lemma 4). 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 4).

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:
\[
\varphi_n^2 = \frac{L}{2\pi \alpha r} h_*^{r-1} \exp \left( -\frac{2\alpha}{h_*^r} \right) = \frac{L}{2\pi \alpha r} \left( \log n \right)^{(1-r)/s} \exp \left( -\frac{2\alpha}{h_*^r} \right) (1 + o(1))
\]
and
\[
\varphi_n^2(\mathbb{L}_2) = L \exp \left( -\frac{2\alpha}{h_*^s} \right).
\]

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 $L_2$-risk bounds

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

(16)

$$
\limsup_{n \to \infty} R_n(L_2, \hat{f}_n, A_{\alpha,r}(L)) \varphi_n^{-2}(L_2) \leq 1,
$$

(17)

where the rates $\varphi_n$ and $\varphi_n(L_2)$ are given in (14) and (15).

The case $r = 1$ and $s = 2$ is of a particular interest. It covers the situation where the noise density $f^\varepsilon$ 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. This case is in the zone $r \leq s/2$, where we get the following behaviour

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

(18)

and

$$
\varphi_n^2(L_2) = \begin{cases} 
L \exp \left( -2\alpha \left( \frac{\log n}{2\beta} \right)^{r/s} \right) (1 + o(1)), & \text{if } r < s/2, \\
L \exp \left( -2\alpha \sqrt{\frac{\log n}{2\beta} + \frac{\alpha^2}{\beta}} \right) (1 + o(1)), & \text{if } r = s/2.
\end{cases}
$$

(19)

The bandwidth (13) depends on the parameters $\alpha, r$ of the class $A_{\alpha,r}(L)$ 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 $L_2$ 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 $\beta$ are supposed to be known since they characterize the known density of noise $f^\varepsilon$.

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

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

(20)
for $n$ large enough so that $\log n/(2\beta) > 1$. Then, for all $(\alpha, L, r) \in \Theta$,

$$
\limsup_{n \to \infty} \sup_{x \in \mathbb{R}} R_n(x, f_n^a, A_{\alpha, r}(L)) \varphi_n^{-2} \leq 1,
$$

and

$$
\limsup_{n \to \infty} R_n(L_2, f_n^a, A_{\alpha, r}(L)) \varphi_n^{-2}(L_2) \leq 1,
$$

where the rates $\varphi_n$ and $\varphi_n(L_2)$ are given 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

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

On the other hand,

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

Therefore, the ratio of the bias term of $f_n^a$ to the variance term of $f_n^a$ both for the pointwise risk and for the $L_2$-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 $\infty$ 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 $f_n^a$ for both risks are asymptotically bounded by $\varphi_n^2$ and $\varphi_n^2(L_2)$ respectively.

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

$$
\frac{L}{2\pi \alpha r} (h_n^a)^{r-1} \exp \left( -2\alpha \left( \frac{\log n}{2\beta} \right)^{r/s} \left[ 1 - \left( \frac{\log n}{2\beta} \right)^{-1/2} \right]^{r/s} \right)
$$

$$
\leq \frac{L(1 + o(1))}{2\pi \alpha r} \left( \frac{\log n}{2\beta} \right)^{(1-r)/s} \exp \left( -2\alpha \left( \frac{\log n}{2\beta} \right)^{r/s} + c \left( \frac{\log n}{2\beta} \right)^{r/s-1/2} \right)
$$

$$
= \varphi_n^2(1 + o(1)),
$$

12
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 $f_n^a$ for the $L_2$-risk is bounded from above by

$$L \exp \left( -2\alpha \left( \frac{\log n}{2\beta} \right)^{r/s} \left[ 1 - \left( \frac{\log n}{2\beta} \right)^{-1/2} \right]^{r/s} \right)$$

$$\leq L \exp \left( -2\alpha \left( \frac{\log n}{2\beta} \right)^{r/s} + c \left( \frac{\log n}{2\beta} \right)^{r/s-1/2} \right) = \varphi_n^2(L_2)(1 + o(1)),$$

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

If $r = s/2$, adaptation to $(\alpha, L)$ 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 $\Phi^c(u) \neq 0, \forall u \in \mathbb{R}$. Let $f_n^a$ be the kernel estimator defined in (2) with bandwidth $h_n = h_n^a$ defined by

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

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 $(\alpha, L) \in \Theta_0$,

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

(21)

$$\limsup_{n \to \infty} R_n(L_2, f_n^a, A_{\alpha,r}(L)) \varphi_n^{-2}(L_2) \leq \exp \left( \frac{\alpha A}{\beta} - \frac{\alpha^2}{\beta^2} \right),$$

(22)

where the rates $\varphi_n$ and $\varphi_n(L_2)$ are given in (18) and (19).

**Proof.** It is easily checked that the bias exponent

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

while for the variance term exponent

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

Since $A > \alpha$, the bias term of $f_n^a$ asymptotically dominates its variance term. Inequalities (21) and (22) now follow from these remarks and the expressions for $\varphi_n^2, \varphi_n(L_2)$ in (18) and (19) with $r = s/2$. \square
Appendix

Lemma 1 For $0 < \alpha, r, L < \infty$,

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

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,

$$
|f(x)| \leq \frac{1}{2\pi} \int |\Phi(u)| du, \quad \forall x \in \mathbb{R}. \tag{23}
$$

By Markov’s inequality

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

Also,

$$
\int |\Phi(u)| I(|\Phi(u)| \exp(2\alpha|u|^r) \leq 1) du \leq 2 \int_0^\infty \exp(-2\alpha u^r) du = 2C(r, \alpha).
$$

Combining the last two inequalities with (23) proves the Lemma. \hfill \Box

Lemma 2 For any positive $\alpha, \beta, r, s$ and for any $A \in \mathbb{R}$ and $B \in \mathbb{R}$, we have

$$
\int_v^\infty u^A \exp(-\alpha u^r) du = \frac{1}{\alpha r} v^{A+1-r} \exp(-\alpha v^r)(1 + o(1)), \quad v \to \infty, \tag{24}
$$

and

$$
\int_0^v u^B \exp(\beta u^s) du = \frac{1}{\beta s} v^{B+1-s} \exp(\beta v^s)(1 + o(1)), \quad v \to \infty. \tag{25}
$$

Proof of this lemma is omitted. It is based on integration by parts and standard evaluations of integrals.

Lemma 3 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

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

for $|x| \geq c_2$.  

14
Proof. From Zolotarev (1986), Th. 2.2.3., formula (2.2.18), we get
\[ p(x) = \frac{|x|^{1/(r-1)}}{2(r-1)} \int_0^1 u(\varphi) \exp(-|x|^{r/(r-1)}u(\varphi))d\varphi, \quad x \neq 0, \quad (26) \]
where
\[ u(\varphi) = \left( \frac{\sin(\pi r\varphi/2)}{\cos(\pi \varphi/2)} \right)^{r/(1-r)} \frac{\cos(\pi(r-1)\varphi/2)}{\cos(\pi\varphi/2)}. \]
Clearly, for \( \varphi \in [1/2, 1] \) we have
\[ 1 \geq \cos(\pi(r-1)\varphi/2) \geq \cos(\pi(r-1)/4) > 0 \]
\[ c_3 \geq \sin(\pi r\varphi/2) \geq c_4 > 0, \]
where \( c_3 > 0 \) and \( c_4 > 0 \) are constants. Thus,
\[ c_6 (\cos(\pi \varphi/2))^{1/(r-1)} \leq u(\varphi) \leq c_5 (\cos(\pi \varphi/2))^{1/(r-1)}, \]
\( \varphi \in [1/2, 1], c_5 > 0, c_6 > 0 \) are constants. Now, if \( \varphi \in [1/2, 1] \)
\[ c_7 (1 - \varphi) \leq \cos(\pi \varphi/2) \leq c_8 (1 - \varphi) \]
for some \( c_7 > 0, c_8 > 0 \). Finally,
\[ c_{10} (1 - \varphi)^{1/(r-1)} \leq u(\varphi) \leq c_9 (1 - \varphi)^{1/(r-1)}, \forall \varphi \in [1/2, 1]. \]
Using (26) and the fact that \( u(\varphi) \geq 0 \) for \( \varphi \in [0, 1] \), we get
\[ p(x) \geq c|x|^{1/(r-1)} \int_{1/2}^1 (1 - \varphi)^{1/(r-1)} \exp(-|x|^{r/(r-1)}c_9(1 - \varphi)^{1/(r-1)}) d\varphi \]
\[ = c|x|^{1/(r-1)} \int_0^{1/2} \varphi^{1/(r-1)} \exp(-c_9(|x|^{r})^{1/(r-1)}) d\varphi. \]
Here and further on \( c > 0 \) are constants, probably different on different occasions.

By change of variables, \( u = (|x|^{r})^{1/(r-1)} \), we get
\[ p(x) \geq c|x|^{1/(r-1)} \int_0^{(|x|^{r/2})^{1/(r-1)}} \frac{u}{|x|^{r/(r-1)}} \exp(-c_9u)u^{r-2} du \]
\[ = c|x|^{-1-r} \int_0^{(|x|^{r/2})^{1/(r-1)}} u^{r-1} \exp(-c_9u) du \]
\[ \geq c|x|^{-1-r} \int_0^{(c_2^{r/2})^{1/(r-1)}} u^{r-1} \exp(-c_9u) du \geq c_1|x|^{-1-r}, \]
for \( |x| \geq c_2 > 0 \). This also implies that \( p(x) > 0, \forall x \neq 0, \)
\[ p(0) = (2\pi)^{-1} \int \exp(-|t|^{r}) dt \neq 0, \]
hence \( p \) is positive on \( \mathbb{R} \). \( \square \)
Lemma 4 Let $0 < r < s < \infty$ and let $h_s = h_s(n)$ be defined by (13), i.e.
\[
\frac{2\alpha}{h_s^r} + \frac{2\beta}{h_s^s} = \log n - (\log \log n)^2.
\]
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
\[
h_s(n) = (\log n/(2\beta))^{-1/s}(1 + o(1)),
\]
\[
h_n^a \exp \left( - \frac{2\alpha}{h_n^r} \right) = h_n^a \exp \left( - \frac{2\alpha}{h_s^r} \right) (1 + o(1)),
\]
\[
h_n^a \exp \left( \frac{2\beta}{h_n^s} \right) = o \left( \exp \left( - \frac{2\alpha}{h_s^r} \right) \right),
\]
for any $a \in \mathbb{R}$, and
\[
h_n^{s+2\gamma-1} \exp \left( \frac{2\beta}{h_n^s} \right) \leq h_n^{s+2\gamma-1} \exp \left( \frac{2\beta}{h_s^s} \right),
\]
for $n$ large enough.

Proof. Define $x_* = h_*^{-s}$, $x_n = h_n^{-s}$, and write, for $t > 0$,
\[
F(t) \overset{\text{def}}{=} 2\beta t + 2\alpha t^{r/s}, \quad F_1(t) \overset{\text{def}}{=} (-b/s) \log t + 2\beta t + 2\alpha t^{r/s}.
\]
Then
\[
F(x_*) = \log n - (\log \log n)^2, \quad F_1(x_n) = \log n + C(1 + o(1)),
\]
for a constant $C \in \mathbb{R}$. We first prove that $x_n$ satisfies
\[
F(x_n) = \log n + C_1 \log \log n(1 + o(1)) + C_2(1 + o(1))
\]
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 (32) 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 (27). Thus $-(b/s)\log x_n = -(b/s)\log \log n(1 + o(1))$, as $n \to \infty$, and write $F(x_n) = F_1(x_n) + (b/s)\log x_n$ to get (33) in view of (32). We have

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

where $a_n = C_1 \log \log n(1 + o(1)) + C_2(1 + o(1)) = O(\log \log n)$, $b_n = (\log \log n)^2$ and $F^{-1}(\cdot)$ is the inverse of $F(\cdot)$. Hence, for some $0 < \tau < 1$ and for $n$ large enough,

$$x_n = F^{-1}(\log n + a_n) = x_* + (F^{-1}(\log n - b_n))'(a_n + b_n) + \frac{1}{2}(F^{-1}(\log n - b_n(1 - \tau) + \tau a_n))''(a_n + b_n)^2. \quad (34)$$

The first and the second derivatives of $F^{-1}$ are given by

$$\begin{align*}
(F^{-1}(y))' &= \frac{1}{F'(F^{-1}(y))} = \frac{1}{2\beta + (2\alpha r/s)(F^{-1}(y))^{r/s-1}}, \\
(F^{-1}(y))'' &= \frac{-(2\alpha r/s)(r/s - 1)(F^{-1}(y))^{r/s-2}}{(2\beta + (2\alpha r/s)(F^{-1}(y))^{r/s-1})^3}.
\end{align*}$$

Hence

$$\frac{1}{2\beta + (2\alpha r/s)x_*^{r/s-1}} = \frac{1}{2\beta} + o(1), \quad n \to \infty. \quad (35)$$

Next, it easy to show that there exists $\bar{y} > 0$ such that

$$y/(4\beta) \leq F^{-1}(y) \leq y/(2\beta) \quad (36)$$

for $y \geq \bar{y}$. Considering $n$ large enough so that $y_n \overset{\text{def}}{=} \log n - b_n(1 - \tau) + a_n\tau \geq \bar{y}$ and using the above expression for $(F^{-1}(y))''$ and (36) we get

$$(F^{-1}(y_n))'' = (F^{-1}(\log n - b_n(1 - \tau) + \tau a_n))'' = O((\log n)^{r/s-2}), \quad n \to \infty.$$

This and (34), (35) imply

$$x_* - x_n = -\frac{1}{2\beta}(1 + o(1))(a_n + b_n) + O\left(\frac{(a_n + b_n)^2}{(\log n)^{2-r/s}}\right) = -\frac{b_n}{2\beta}(1 + o(1)). \quad (37)$$

Using this representation we obtain

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

17
since \( b_n = (\log \log n)^2, \ x_* = (2\beta)^{-1} \log n \ (1 + o(1)) \) and \( r < s \). This and the fact that \((h_n/h_\ast)^a = (x_*/x_n)^a/s = 1 + o(1)\) imply (28). Next, (29) follows directly from the definition of \( h_\ast \) and from (27). To prove (30), note that, in view of (37),

\[
\frac{h_\ast^{s+2\gamma-1}}{h_n^{s+2\gamma-1}} \exp\left(\frac{2\beta}{h_\ast} - \frac{2\beta}{h_n}\right) = (1 + o(1)) \exp(2\beta(x_* - x_n))
\]

for \( n \) large enough. \( \square \)

**Acknowledgement.** The results of this paper were presented at the Conference “Rencontres de statistiques mathématiques”, CIRM Luminy, 2001. Later, Fabienne Comte and Marie-Luce Taupin suggested a different estimator for the same problem refraining from studying the optimality of rates issue (Comte and Taupin (2003)). We would like to thank them for discussion of the results.

**References**


---

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

**2** Modal’X, Université Paris X 200, avenue de la République 92001 Nanterre Cedex, France E-mail: butucea@ccr.jussieu.fr

---

19