

Butucea, Cristina

Working Paper

Two adaptive rates of convergence in pointwise density estimation

SFB 373 Discussion Paper, No. 1999,20

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Butucea, Cristina (1999) : Two adaptive rates of convergence in pointwise density estimation, SFB 373 Discussion Paper, No. 1999,20, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10056124>

This Version is available at:

<https://hdl.handle.net/10419/61774>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Two adaptive rates of convergence in pointwise density estimation

Cristina BUTUCEA

Humboldt-Universität zu Berlin, SFB373, Spandauer Strasse 1

D 10117 Berlin, Germany

Université Paris VI, UPRES - A 7055 CNRS, 4, Place Jussieu,

F 75005 Paris, France

(E-mail: butucea@wiwi.hu-berlin.de, butucea@ccr.jussieu.fr)

Abstract

We consider density pointwise estimation and look for best attainable asymptotic rates of convergence. The problem is adaptive, which means that the regularity parameter, β , describing the class of densities, varies in a set B . We shall consider, successively, two classes of densities, issued from a generalization of \mathbb{L}_2 Sobolev classes: $W(\beta, p, L)$ and $M(\beta, p, L)$.

Keywords: nonparametric density estimation, adaptive rates, Sobolev classes

1 Introduction

1.1 Adaptivity

We want to estimate the common probability density $f : \mathbb{R} \rightarrow [0, +\infty)$ of n independent, identically distributed random variables X_1, \dots, X_n , at a real point x_0 . We assume that f belongs to a large nonparametric class of functions, $H_\beta = H(\beta, p, L)$, characterized by its smoothness (e.g., order of derivability), β , a norm \mathbb{L}_p , $p > 1$ and a radius $L > 0$.

For any estimator \hat{f}_n of f , fixed x_0 real and $q > 1$ we consider a sequence $\varphi_{n,\beta}$ of positive numbers and define the **maximal risk** over the class H_β :

$$R_{n,\beta}(\hat{f}_n, \varphi_{n,\beta}, H_\beta) = \sup_{f \in H_\beta} \varphi_{n,\beta}^{-q} E_f \left[\left| \hat{f}_n(x_0) - f(x_0) \right|^q \right], \quad (1.1)$$

where $E_f(\cdot)$ is the expectation with respect to the distribution P_f of X_1, \dots, X_n , when the underlying probability density is f .

We say that $\varphi_{n,\beta}$ is an **optimal rate of convergence** over the class $H_\beta = H(\beta, p, L)$, if the maximal risk over this class stays positive, for all possible estimators, asymptotically and if there is an estimator whose maximal risk stays finite asymptotically. Minimax theory is concerned with finding the estimators attaining the optimal rates, which are given by minimizing the maximal risk, over all estimators.

We are interested in adaptive estimation, which means that the regularity parameter β is supposed unknown within a given set. An estimator \hat{f}_n is called **optimal rate adaptive**

if, for the optimal rate of convergence over that class, $\varphi_{n,\beta}$, and a constant $C > 0$, we have

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta} \left(\widehat{f}_n, \varphi_{n,\beta}, H_\beta \right) \leq C < \infty, \quad (1.2)$$

where B is a non-empty set of values.

We shall prove here, that over two different classes of probability density functions, to be defined below, commonly denoted by $H_\beta = H(\beta, p, L)$, we can find no optimal rate adaptive estimator. Similar results were obtained by Lepskii [20], Brown and Low [3] (on Hölder classes of functions) and Tsybakov [27] (on \mathbb{L}_2 Sobolev classes), for the Gaussian white noise model. They are characteristic for the pointwise (not global) estimation. We shall then introduce the definition of adaptive rate of convergence, which is a modification by Tsybakov [27] of the definition of Lepskii [20] (see also Lepskii [21] and [22]). We also compute the adaptive rate over the same classes of functions as well as the corresponding rate adaptive estimators.

More precisely, let us define the considered classes of densities. At first, we define for integer $\beta \geq 1$, $L > 0$ and $1 < p < \infty$ the class of functions in \mathbb{L}_p

$$W(\beta, p, L) = \left\{ f : \mathbb{R} \rightarrow [0, +\infty) : \int_{\mathbb{R}} f = 1, \int_{\mathbb{R}} \left| f^{(\beta)}(x) \right|^p dx \leq L^p \right\},$$

where $f^{(\beta)}$, the derivative of order β of f , is supposed to exist.

Secondly, let us introduce for any absolutely integrable function $f : \mathbb{R} \rightarrow [0, +\infty)$ its Fourier transform $\mathcal{F}(f)(x) = \int_{\mathbb{R}} f(y) e^{-ixy} dy$, for any x in \mathbb{R} . We define now for real $\beta > 1 - \frac{1}{p}$ and $2 \leq p < \infty$ the class of absolutely integrable functions whose Fourier transforms belong to \mathbb{L}_p and

$$M(\beta, p, L) = \left\{ f : \mathbb{R} \rightarrow [0, +\infty) : \int_{\mathbb{R}} f = 1, \int_{\mathbb{R}} |\mathcal{F}(f)(x)|^p |x|^{p\beta} dx \leq L^p \right\}.$$

From the results of optimal recovery of Donoho and Low [9], it is straightforward to obtain the optimal pointwise rates of convergence over these classes:

$$\varphi_{n,\beta}(W(\beta, p, L)) = \left(\frac{1}{n} \right)^{\frac{\beta-1/p}{2(\beta-1/p)+1}} \quad \text{and} \quad \varphi_{n,\beta}(M(\beta, p, L)) = \left(\frac{1}{n} \right)^{\frac{\beta+1/p-1}{2(\beta+1/p)-1}}. \quad (1.3)$$

In this paper, we prove that no optimal adaptive estimator can be found and we look for the adaptive rates of convergence on previously defined classes $W(\beta, p, L)$ and $M(\beta, p, L)$, for β belonging to a set B_{N_n} to be defined for each class. We prove that the adaptive rates of convergence are within a factor $\log n$ slower than the optimal rates.

Remark 1.1 *If p' is the conjugate of p (i.e. $1/p + 1/p' = 1$), then the optimal rates of convergence (1.3) coincide for integer β , on the classes $W(\beta, p, L)$ and $M(\beta, p', L)$ (as well as the adaptive rates (2.5) below). Moreover, by a result of Stein and Weiss [24] we have that for integer $\beta \geq 1$ and $1 < p \leq 2$*

$$M(\beta, p', L) \subseteq W(\beta, p, L).$$

Thus, parts of our results on a scale of classes can be deduced from the results on the other scale, for certain values of the parameters. Nevertheless, our setups are considerably larger and the classes W and M are not compatible except in the particular case described above. For these reasons, we prefer the above notation and give independent proofs for both setups.

1.2 Previous results

The asymptotic study of minimax risks of estimators in the nonparametric framework, was developed considerably since the first results of Stone [25] and [26], Bretagnolle et Huber [2], Ibragimov et Hasminskii [15] and [16]. Beside the density model, nonparametric regression and Gaussian white noise models were studied. Estimation was done over Hölder, Sobolev or Besov classes. For an overview of the results in this area see Korostelev and Tsybakov [19] and Härdle, Kerkyacharian, Picard and Tsybakov [14].

Almost optimal rates of convergence in density pointwise estimation over \mathbb{L}_p Sobolev classes, $W(\beta, p, L)$, were obtained by Wahba [28], where technics of Farrel [11] for the proof of the lower bounds issued a rate of $(1/n)^{\frac{\beta-1/s}{2(\beta-1/s)+1}}$, for $s = p + \varepsilon$, $\varepsilon > 0$ arbitrary small. Note that the optimal rate for $W(\beta, p, L)$, as noted in (1.3), is given by this expression with $\varepsilon = 0$.

Technics of optimal recovery of Donoho [5], Donoho and Liu [8], Donoho and Low [9] allow to compute optimal rates of convergence for different risks, in different setups. In these papers the classes $M(\beta, p, L)$ and the corresponding rate in (1.3) first appear.

Lepskii [20], Brown and Low [3] showed that for pointwise estimation on the Hölder classes optimal rate adaptive estimators can not be found, both in Gaussian white noise and density models. In the Gaussian white noise model, Lepskii [20] first considered the problem of finding the adaptive rates. He showed that a loss of logarithmic order is unavoidable and introduced a procedure providing the adaptive estimator. For a detailed overview of adaptive rates of convergence we refer to Donoho, Johnstone, Kerkyacharian, Picard [6], Härdle, Kerkyacharian, Picard, Tsybakov [14] who give adaptive rates over Besov classes using the wavelet thresholding procedure. Lepski, Mammen and Spokoiny [23], Goldenshluger and Nemirovski [12], Juditsky [17] gave also adaptive rates of convergence using Lepski's scheme of adaptation. Most of these results are obtained for the Gaussian white noise model.

In density estimation, wavelet techniques were used in the minimax adaptive setup for Besov classes and \mathbb{L}_p risk, by Donoho, Johnstone, Kerkyacharian, Picard [7], Kerkyacharian, Picard and Tribouley [18] and Juditsky [17]. Sharp results, where the asymptotic value of the maximal risk was found explicitly, were obtained over \mathbb{L}_2 Sobolev classes in \mathbb{L}_2 risk by Efromovich [10] and Golubev [13] and pointwise in Butucea [4].

In this paper, we are interested in adaptive rates in pointwise density estimation over \mathbb{L}_p Sobolev classes, $W(\beta, p, L)$ and $M(\beta, p, L)$.

2 Results

We consider adaptive density estimation problem, at fixed real point x_0 , over the classes $H_\beta = H(\beta, p, L)$, when β belongs to the discrete set $B_{N_n} = \{\beta_1, \dots, \beta_{N_n}\}$.

Assumption (A) The set B_{N_n} is such that $\beta_1 < \dots < \beta_{N_n} < \infty$, for a non-decreasing sequence of positive integers N_n . From now on, we shall consider two setups. When $H_\beta = W(\beta, p, L)$ the set B_{N_n} contains positive integer values of β ($\beta_1 \geq 1$) and $p > 1$, while $H_\beta = M(\beta, p, L)$ implies that β can take real values, ($\beta_1 > 1 - 1/p$) and $p \geq 2$. Moreover, we suppose that $\lim_{n \rightarrow \infty} \beta_{N_n} = \infty$ and if $\Delta_n = \min_{i=1..N_n-1} |\beta_{i+1} - \beta_i|$

we assume that it satisfies

$$\limsup_{n \rightarrow \infty} \Delta_n < +\infty \quad (2.1)$$

together with

$$\lim_{n \rightarrow \infty} \frac{\Delta_n \log n}{\beta_{N_n}^2 \log \log n} = \infty. \quad (2.2)$$

The following definition of an adaptive rate of convergence was introduced by Lepski, see Tsybakov [27]. The original definition of adaptive rate of convergence by Lepskii [20] is not used here since it has a more special form.

Definition 2.1 *The sequence $\psi_{n,\beta}$ is an **adaptive rate of convergence** over the scale of classes $\{H_\beta, \beta \in B_{N_n}\}$, if*

1. *There exists an estimator f_n^* , independent of β over B_{N_n} , which is called **rate adaptive estimator**, such that*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B_{N_n}} R_{n,\beta}(f_n^*, \psi_{n,\beta}, H_\beta) < \infty, \quad (2.3)$$

2. *If there exist another sequence of positive reals $\rho_{n,\beta}$ and an estimator f_n^{**} such that*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B_{N_n}} R_{n,\beta}(f_n^{**}, \rho_{n,\beta}, H_\beta) < \infty$$

and, at some β' in B_{N_n} , $\frac{\rho_{n,\beta'}}{\psi_{n,\beta'}} \xrightarrow{n \rightarrow \infty} 0$, then there is another β'' in B_{N_n} such that $\frac{\rho_{n,\beta'}}{\psi_{n,\beta'}} \cdot \frac{\rho_{n,\beta''}}{\psi_{n,\beta''}} \xrightarrow{n \rightarrow \infty} +\infty$.

Note that condition (2.3) introduces a wide class of rates. We choose between those rates by a criterion of uniformity over the set B_{N_n} , expressed in the second part of Definition 2.1. If some other rate satisfies a condition similar to (2.3) and if this rate is faster at some point β' then the loss at some other point β'' has to be infinitely greater for large sample sizes n .

Remark 2.2 *If an optimal adaptive estimator exists, it is also rate adaptive.*

Indeed, an optimal adaptive estimator satisfies (2.3) by definition, for the optimal rate of convergence $\psi_{n,\beta} = \varphi_{n,\beta}$. We can easily verify that in this case condition 2 in Definition 2.1 is redundant, since such a sequence $\rho_{n,\beta}$ can not exist.

In what follows we assign to any β in B_{N_n} the value

$$\tilde{\beta} = \tilde{\beta}(H) = \begin{cases} \beta - 1/p + 1/2, & \text{if } H = W \\ \beta + 1/p - 1/2, & \text{if } H = M \end{cases}, \quad (2.4)$$

where equalities $H = W$ and $H = M$ denote the cases when we consider the scales of classes $\{W(\beta, p, L), \beta \in B_{N_n}\}$ or $\{M(\beta, p, L), \beta \in B_{N_n}\}$, respectively. We remark that in both setups: $\tilde{\beta} > 1/2$.

Let us define $B_- = B_{N_n} \setminus \{\beta_{N_n}\}$ and

$$\psi_{n,\beta} = \psi_{n,\beta}(H_\beta) = \begin{cases} (\log n/n)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}, & \text{if } \beta \in B_- \\ (1/n)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}, & \text{if } \beta = \beta_{N_n} \end{cases}. \quad (2.5)$$

Then the rate $\psi_{n,\beta}(H_\beta)$ is slower than the optimal rate of convergence, except for the last point β_{N_n} . As by our hypothesis $\lim_{n \rightarrow \infty} \beta_{N_n} = \infty$, this asymptotic phenomenon is not characteristic and we can use the set B_- instead of B_{N_n} .

2.1 The adaptive procedure

Let us proceed to the construction of the estimator f_n^* called adaptive estimator. We start for each β in B_{N_n} with the corresponding kernel estimator

$$f_{n,\beta}(x_0) = \frac{1}{nh_{n,\beta}} \sum_{i=1}^n K_\beta \left(\frac{X_i - x_0}{h_{n,\beta}} \right). \quad (2.6)$$

Here the kernel K_β is defined in the next section (differently for each setup) and the bandwidth is in both problems

$$h_{n,\beta} = \left(\frac{\log n}{n} \right)^{\frac{1}{2\tilde{\beta}}}, \text{ if } \beta \in B_- \text{ and } h_{n,\beta_{N_n}} = \left(\frac{1}{n} \right)^{\frac{1}{2\tilde{\beta}_{N_n}}},$$

where $\tilde{\beta} = \tilde{\beta}(H_\beta)$ and $\tilde{\beta}_{N_n} = \tilde{\beta}_{N_n}(H_{\beta_{N_n}})$ in (2.4). We shall evaluate the regularity β of the estimated density and replace it into the kernel estimator $f_{n,\beta}$ in order to obtain f_n^* , the adaptive estimator, in the spirit of Lepskii [20].

More precisely, let $a > 0$ be a sufficiently large constant and

$$\eta_{n,\beta} = a \left(\frac{\log n}{n} \right)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}.$$

Then, we define

$$\hat{\beta} = \hat{\beta}(H_\beta) = \max \{ \beta \in B_{N_n} : |f_{n,\beta}(x_0) - f_{n,\gamma}(x_0)| \leq \eta_{n,\gamma}, \forall \gamma < \beta, \gamma \in B_{N_n} \}.$$

In the sequel, $\tilde{\gamma}$ (appearing in $\eta_{n,\gamma}$) is defined as in (2.4). Finally,

$$f_n^*(x_0) = f_{n,\hat{\beta}}(x_0). \quad (2.7)$$

2.2 Statement of results

Theorem 2.3 *In both pointwise density estimation problems described above, we can find no optimal rate adaptive estimators (see Definition 1.2) over the scale of classes $\{H(\beta, p, L), \beta \in B\}$, as soon as B has at least two different elements and $B \subseteq B_{N_n}$, where B_{N_n} satisfies **Assumption (A)**.*

Theorem 2.4 *The estimator $f_n^*(x_0)$ of $f(x_0)$, in (2.7), is rate adaptive estimator and $\psi_{n,\beta}(H_\beta)$ in (2.5) is the adaptive rate of convergence in the sense of Definition 2.1, over the scale $\{H(\beta, p, L), \beta \in B_{N_n}\}$, where the set B_{N_n} satisfies **Assumption (A)**.*

The proof is organized as follows. In Section 3 we prove that $f_n^*(x_0)$ in (2.7) satisfies, for a constant $C > 0$,

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B_{N_n}} R_{n,\beta}(f_n^*, \psi_{n,\beta}, H_\beta) \leq C < \infty. \quad (2.8)$$

This result will be called the upper bound. Section 4 is devoted to the proof of the lower bound

$$\liminf_{n \rightarrow \infty} \inf_{f_n} \sup_{\alpha \in \{\gamma, \beta\}} R_{n,\alpha}(f_n, \psi_{n,\alpha}, H_\alpha) \geq c > 0,$$

where γ and β are in B_{N_n} , arbitrary chosen elements such that $\gamma < \beta$, $c > 0$ and the infimum is taken over all possible estimators f_n of f . These relations, Theorem 2.3 (proved in Section 5) and the fact that $\psi_{n,\beta}(H_\beta)$ is the adaptive rate of convergence over the set B_{N_n} (see also Section 5) imply Theorem 2.4.

3 Upper bounds

We shall prove that the estimator f_n^* , independent of β in B_{N_n} , defined in (2.7), is such that the upper bound (2.8) holds. Throughout this section, C , c_i and C_i , $i = 1, 2, \dots$, denote positive constants, depending possibly on fixed q , β_1 and L .

3.1 Auxiliary results

Definition 3.1 *Let the density f belong to the class $H_\beta = H(\beta, p, L)$. Define for any kernel estimator $f_{n,\gamma}$ of f (see (2.6)), with γ, β in B_{N_n} such that $\gamma \leq \beta$ its bias term*

$$B_{n,\gamma} = B_{n,\gamma}(x_0, H_\beta) = |E_f[f_{n,\gamma}(x_0)] - f(x_0)|,$$

and its stochastic term

$$Z_{n,\gamma} = Z_{n,\gamma}(x_0, H_\beta) = |f_{n,\gamma}(x_0) - E_f[f_{n,\gamma}(x_0)]|.$$

Besov, Il'in and Nikol'skii [1], Theorem 15.1 implies the following:

Lemma 3.2 *Let β, γ be integers and $0 \leq \gamma < \beta$, $1 \leq p_0, p_1, p \leq \infty$, $\beta > 1/p$. If there exists $\theta \in (\gamma/\beta, 1)$ such that*

$$\frac{1}{p_0} - \gamma = (1 - \theta) \frac{1}{p_1} + \theta \left(\frac{1}{p} - \beta \right), \quad (3.1)$$

then any function $f \in \mathbb{L}_{\beta_1}(\mathbb{R})$ with $\|f^{(\beta)}\|_p < \infty$ satisfies

$$\|f^{(\gamma)}\|_{p_0} \leq C \|f\|_{p_1}^{1-\theta} \|f^{(\beta)}\|_p^\theta,$$

where C is a constant that depends only on $p_0, p_1, p, \beta, \gamma$.

Lemma 3.3 *There exists a finite constant Λ depending on L , β and p only such that*

$$\sup_{f \in H(\beta, p, L)} \|f\|_\infty \leq \Lambda.$$

Proof. For $f \in W(\beta, p, L)$, we apply the previous result with $\gamma = 0$, $p_0 = \infty$, $p_1 = 1$. Then (3.1) takes the form

$$0 = (1 - \theta) + \theta \left(\frac{1}{p} - \beta \right),$$

which implies $\theta = 1/(\beta + 1 - 1/p)$. Then $\theta \in (\gamma/\beta, 1)$ if $\beta > 1/p$ which holds by hypothesis. Thus, we apply the previous result, Lemma 3.2, and get

$$\|f\|_\infty \leq C \|f\|_1^{1-\theta} \left\| f^{(\beta)} \right\|_p^\theta \leq CL^\theta,$$

for all f in $W(\beta, p, L)$.

If $f \in M(\beta, p, L)$, then $\|\mathcal{F}(f)\|_\infty \leq 1$ since f is a density. We have

$$\begin{aligned} \|f\|_\infty &\leq \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{F}(f)(y)| dy \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{F}(f)(y)| \left(1 + |y|^\beta\right) \frac{dy}{1 + |y|^\beta} \\ &\leq \frac{1}{2\pi} \left(\int_{\mathbb{R}} |\mathcal{F}(f)(y)|^p \left(1 + |y|^\beta\right)^p dy \right)^{1/p} \left(\int_{\mathbb{R}} \frac{dy}{\left(1 + |y|^\beta\right)^{p'}} \right)^{1/p'} \end{aligned}$$

where $1/p + 1/p' = 1$. This is less than a constant $\Lambda(L, \beta, p) > 0$, for f in the class $M(\beta, p, L)$. \square

Lemma 3.4 *If $f \in H(\beta, p, L)$ and γ is in B_{N_n} such that $\gamma < \beta$, then $f \in H(\gamma, p, L')$, where $L' > 0$ depends only upon L and p .*

Proof. For classes $W(\beta, p, L)$ put $p_0 = p$, $p_1 = 1$ in the auxiliary Lemma 3.2. Then (3.1) takes form

$$\frac{1}{p} - \gamma = (1 - \tilde{\theta}) + \tilde{\theta} \left(\frac{1}{p} - \beta \right),$$

which gives $\tilde{\theta} = (\gamma + 1 - 1/p) / (\beta + 1 - 1/p)$ and thus $\tilde{\theta} \in (\gamma/\beta, 1)$ if $\beta > 1/p$ (true, by hypothesis). By Lemma 3.2 we get

$$\left\| f^{(\gamma)} \right\|_p \leq \tilde{C} \|f\|_1^{1-\tilde{\theta}} \left\| f^{(\beta)} \right\|_p^{\tilde{\theta}} \leq \tilde{C} L^{\tilde{\theta}},$$

for all f in $W(\beta, p, L)$.

For $f \in M(\beta, p, L)$, as $p\beta > 1$ and $\|\mathcal{F}(f)\|_\infty \leq 1$, we write

$$\begin{aligned} \int_{\mathbb{R}} |\mathcal{F}(f)(y)|^p |y|^{p\gamma} dy &\leq \int_{|y| \leq 1} |\mathcal{F}(f)(y)|^p dy + \int_{|y| > 1} |\mathcal{F}(f)(y)|^p |y|^{p\beta} dy \\ &\leq 1 + L^p. \end{aligned}$$

□

Lemma 3.5 *If γ and β are in B_{N_n} such that $\gamma \leq \beta$ and if f belongs to $H_\beta = H(\beta, p, L)$ then there exists $b_\gamma(H_\beta) > 0$ (given in the proof and depending also on L and p), such that*

$$\begin{aligned} B_{n,\gamma}(x_0, H_\beta) &\leq b_\gamma(H_\beta) h_{n,\gamma}^{\gamma-1/p}, \text{ if } H_\beta = W(\beta, p, L), \\ B_{n,\gamma}(x_0, H_\beta) &\leq b_\gamma(H_\beta) h_{n,\gamma}^{\gamma-1+1/p}, \text{ if } H_\beta = M(\beta, p, L), \end{aligned}$$

and

$$E_f [Z_{n,\gamma}(x_0, H_\beta)]^2 \leq \frac{\Lambda \|K_\gamma\|_2^2}{nh_{n,\gamma}} \stackrel{Def}{=} s_{n,\gamma}^2. \quad (3.2)$$

Moreover, for the kernels $\{K_\beta, \beta \in B_{N_n}\}$ used in the proof, we can find constants K_{\max} , k_{\min} , k_{\max} and b_{\max} depending possibly on fixed p and β_1 , such that

$$\|K_\beta\|_\infty \leq K_{\max}, k_{\min} \leq \|K_\beta\|_2 \leq k_{\max}$$

for all β in B_{N_n} and $b_\gamma(H_\beta) \leq b_{\max}$ for all γ and β in B_{N_n} such that $\gamma \leq \beta$.

Remark 3.6 *From now on, $\tilde{\gamma} = \tilde{\gamma}(H_\beta)$ is obtained as in (2.4). Then Lemma 3.5 says that*

$$B_{n,\gamma}(x_0, H_\beta) \leq b_\gamma h_{n,\gamma}^{\tilde{\gamma}(H_\beta)-1/2}.$$

Proof. If $H_\beta = W_\beta = W(\beta, p, L)$, let us introduce a kernel K_γ of order γ , in the expression of the kernel estimator (2.6). Such a kernel must be bounded uniformly in γ ($\|K_\gamma\|_\infty \leq K_{\max}$, for all γ in B_{N_n}), absolutely integrable, with a bounded \mathbb{L}_2 norm ($k_{\min} \leq \|K_\gamma\|_2 \leq k_{\max}$, for all γ in B_{N_n}), such that $\int_{\mathbb{R}} K_\gamma(y) dy = 1$, $\int_{\mathbb{R}} y^j K_\gamma(y) dy = 0$ for $j = 1, \dots, \gamma - 1$ and

$$\int_{\mathbb{R}} |K_\gamma(y)| |y|^{\gamma-1/p} dy \leq L_0 < \infty, \quad (3.3)$$

where L_0 depends only on fixed p and β_1 . It is not difficult to find examples of such kernels. For example, the kernel K_γ having Fourier transform $\mathcal{F}(K_\gamma)(u) = 1/(1 + |u|^{p\gamma})$ satisfy these conditions and the proofs are given later on.

From now on we denote $f = \int_{\mathbb{R}}$. Then the bias can be bounded as follows

$$\begin{aligned} B_{n,\gamma}(x, W_\beta) &= \left| \int K_\gamma(y) [f(x + yh_{n,\gamma}) - f(x)] dy \right| \\ &\leq \left| \int K_\gamma(y) \sum_{j=1}^{\gamma-1} \frac{(yh_{n,\gamma})^j}{j!} f^{(j)}(x) dy \right| \\ &\quad + \left| \int K_\gamma(y) \int_x^{x+yh_{n,\gamma}} \frac{(x+yh_{n,\gamma}-u)^{\gamma-1}}{(\gamma-1)!} f^{(\gamma)}(u) du dy \right| \\ &\leq \left| \sum_{j=1}^{\gamma-1} \frac{h_{n,\gamma}^j}{j!} f^{(j)}(x) \int y^j K_\gamma(y) dy \right| + \\ &\quad + \int |K_\gamma(y)| \left\| f^{(\gamma)} \right\|_p \frac{|yh_{n,\gamma}|^{\gamma-1/p}}{(\gamma-1)! ((\gamma-1)p' + 1)^{1/p'}} dy, \end{aligned}$$

where the first term is zero by hypotheses on the kernel and we applied the Hölder inequality with $1/p + 1/p' = 1$ for the second term. This gives

$$\begin{aligned} B_{n,\gamma}(x, W_\beta) &\leq \frac{L'}{(\gamma-1)!} \frac{h_{n,\gamma}^{\gamma-1/p}}{((\gamma-1)p' + 1)^{1/p'}} \int |K_\gamma(y)| |y|^{\gamma-1/p} dy \\ &\leq b_\gamma(W_\beta) h_{n,\gamma}^{\tilde{\gamma}(W_\beta)-1/2}, \end{aligned}$$

where

$$b_\gamma(W_\beta) = \frac{L'}{(\gamma-1)!} \frac{\int |K_\gamma(y)| |y|^{\gamma-1/p} dy}{((\gamma-1)p' + 1)^{1/p'}}.$$

We can also see that $b_\gamma(W_\beta) \leq b_{\max}$, b_{\max} depending only on p , L and β_1 , for all γ and β in B_{N_n} , $\gamma \leq \beta$.

If $H_\beta = M_\beta = M(\beta, p, L)$, let us choose the kernel K_γ defined by its Fourier transform as follows

$$\mathcal{F}(K_\gamma)(u) = \frac{1}{1 + |u|^{p\gamma}}.$$

This kernel has, by Plancherel's formula:

$$\begin{aligned} \|K_\gamma\|_2 &= \frac{1}{\sqrt{2\pi}} \|\mathcal{F}(K_\gamma)\|_2 = \frac{1}{\sqrt{2\pi}} \int \frac{du}{(1 + |u|^{p\gamma})^2} \\ &\geq \frac{1}{\sqrt{2\pi}} \int_{|u| \leq 1} \frac{du}{(1 + |u|^{p\beta_1})^2} = k_{\min}(p, \beta_1), \end{aligned}$$

also

$$\|K_\gamma\|_2 \leq 1 + \frac{1}{\sqrt{2\pi}} \int_{|u| > 1} \frac{du}{(1 + |u|^{p\beta_1})^2} = k_{\max}(p, \beta_1)$$

and

$$\|K_\gamma\|_\infty \leq \frac{1}{2\pi} \int |\mathcal{F}(K_\gamma)(u)| du \leq 1 + \frac{1}{2\pi} \int_{|u| > 1} \frac{du}{1 + |u|^{p\beta_1}} = K_{\max}(p, \beta_1),$$

since $p\beta_1 > 1$, in our setting. Then the bias is

$$\begin{aligned} B_{n,\gamma}(x, M_\beta) &= \left| \int \frac{1}{h_{n,\gamma}} K_\gamma\left(\frac{y-x}{h_{n,\gamma}}\right) f(y) dy - f(x) \right| \\ &= \frac{1}{2\pi} \left| \int \mathcal{F}(f)(y) e^{ixy} [\mathcal{F}(K_\gamma)(h_{n,\gamma}y) - 1] dy \right| \\ &\leq \frac{1}{2\pi} \int |\mathcal{F}(f)(y)| \frac{|h_{n,\gamma}y|^{p\gamma}}{1 + |h_{n,\gamma}y|^{p\gamma}} dy. \end{aligned}$$

Then we apply Hölder's inequality for $1/p + 1/p' = 1$ as follows

$$\begin{aligned} B_{n,\gamma}(x, M_\beta) &\leq \frac{h_{n,\gamma}^\gamma}{2\pi} \int |\mathcal{F}(f)(y)| |y|^\gamma \frac{|h_{n,\gamma} y|^{(p-1)\gamma}}{1 + |h_{n,\gamma} y|^{p\gamma}} dy \\ &\leq \frac{L' h_{n,\gamma}^{\gamma-1/p'}}{2\pi} \left(\int \frac{|y|^{p\gamma}}{(1 + |y|^{p\gamma})^{p'}} dy \right)^{1/p'} = b_\gamma(M_\beta) h_{n,\gamma}^{\tilde{\gamma}(M)-1/2}, \end{aligned}$$

where L' is the constant from Lemma 3.4 and

$$b_\gamma(M_\beta) = \frac{L'}{2\pi} \left(\int \frac{|y|^{p\gamma}}{(1 + |y|^{p\gamma})^{p'}} dy \right)^{1/p'}.$$

This term is bounded as follows

$$b_\gamma(M_\beta) \leq \frac{L'}{2\pi} \left(\int_{|y|\leq 1} \frac{dy}{(1 + |y|^{p\beta_1})^{p'}} + \int_{|y|>1} \frac{dy}{|y|^{p'\beta_1}} \right)^{1/p'} = b_{\max}(p, L, \beta_1).$$

Let us check at last that condition (3.3) is fulfilled:

$$\begin{aligned} \int_{\mathbb{R}} |K_\gamma(y)| |y|^{\gamma-1/p} dy &\leq \int_{|y|\leq 1} |K_\gamma(y)| dy + \int_{|y|>1} |K_\gamma(y)| |y|^{\gamma-1/p} dy \\ &\leq K_{\max} + \left(\int_{|y|>1} |K_\gamma(y)|^2 |y|^{2\gamma} dy \right)^{1/2} \left(\int_{|y|>1} |y|^{-2/p} dy \right)^{1/2} \\ &\leq L_0(p, \beta_1). \end{aligned}$$

For the variance term, we write, using Lemma 3.3

$$E_f [Z_{n,\gamma}(x, H_\beta)]^2 \leq \frac{1}{nh_{n,\gamma}} \int \frac{1}{h_{n,\gamma}} K_\gamma^2 \left(\frac{y-x}{h_{n,\gamma}} \right) f(x) dx \leq \frac{\Lambda \|K_\gamma\|_2^2}{nh_{n,\gamma}}.$$

□

Let us recall the following inequalities (see e.g. Härdle, Kerkyacharian, Picard, Tsybakov [14]).

Lemma 3.7 Rosenthal's inequality: *Let $q \geq 2$ and Y_1, \dots, Y_n be independent random variables such that $E[Y_i] = 0$, $E[|Y_i|^q] < \infty$. Then there exists $C(q)$ a constant depending on q such that*

$$E \left[\left| \sum_{i=1}^n Y_i \right|^q \right] \leq C(q) \left\{ \sum_{i=1}^n E[|Y_i|^q] + \left(\sum_{i=1}^n E[Y_i^2] \right)^{q/2} \right\}.$$

Bernstein's inequality: *Let Y_1, \dots, Y_n be i.i.d. random variables such that $|Y_i| \leq M$, $E[Y_i] = 0$ and denote $b_n^2 = \sum_{i=1}^n E[Y_i^2]$. Then for any $\lambda > 0$,*

$$P \left[\left| \sum_{i=1}^n Y_i \right| \geq \lambda \right] \leq 2 \exp \left\{ - \frac{\lambda^2}{2(b_n^2 + \lambda M/3)} \right\}.$$

Lemma 3.8 *If f belongs to $H_\beta = H(\beta, p, L)$ and $\gamma < \beta$, if K_γ is the kernel function and*

$$Z_{n,\gamma}(x_0, H_\beta) = \left| \frac{1}{nh_{n,\gamma}} \sum_{i=1}^n \left[K_\gamma \left(\frac{X_i - x_0}{h_{n,\gamma}} \right) - E_f K_\gamma \left(\frac{X_i - x_0}{h_{n,\gamma}} \right) \right] \right|$$

then for any $u > 0$

$$P_f [Z_{n,\gamma}(x_0, H_\beta) \geq u] \leq 2 \exp \left\{ -\frac{u^2}{2s_{n,\gamma}^2(1+c_0u)} \right\},$$

where $c_0 > 0$ does not depend on γ .

Proof. Indeed, we can apply Bernstein's inequality for $\lambda = nu$ and the i.i.d., centered variables

$$Y_i = \frac{1}{h_{n,\gamma}} \left[K_\gamma \left(\frac{X_i - x_0}{h_{n,\gamma}} \right) - E_f K_\gamma \left(\frac{X_i - x_0}{h_{n,\gamma}} \right) \right],$$

bounded as follows: $|Y_i| \leq 2 \|K_\gamma\|_\infty / h_{n,\gamma}$. Then $b_n^2 \leq s_{n,\gamma}^2 = \Lambda \|K_\gamma\|_2^2 / (nh_{n,\gamma})$ by (3.2) and, by Lemma 3.5, $2 \|K_\gamma\|_\infty / (\Lambda \|K_\gamma\|_2^2) \leq 2K_{\max} / (\Lambda k_{\min}^2) = c_0$, which does not depend on γ .

□

Remark 3.9 *For $q > 1$, we can find a constant $c(q) > 0$ such that the stochastic term of the kernel estimator satisfies*

$$E_f [Z_{n,\gamma}(x_0, H_\beta)]^q \leq c(q) s_{n,\gamma}^q,$$

where we denoted $s_{n,\gamma}^2 = (\Lambda \|K_\gamma\|_2^2) / (nh_{n,\gamma})$.

Indeed, for $q > 2$, we apply Rosenthal's inequality to the previous centered variables Y_i , bounded as follows: $|Y_i| \leq 2 \|K_\gamma\|_\infty / h_{n,\gamma}$. Then we can find a constant depending on q , $c'(q)$, such that

$$E_f \left[\left| \frac{1}{n} \sum_{i=1}^n Y_i \right| \right]^q \leq c'(q) \left\{ \left(\frac{2 \|K_\gamma\|_\infty}{nh_{n,\gamma}} \right)^{q-2} \frac{1}{n} E_f Y_1^2 + \left(\frac{1}{n} E_f Y_1^2 \right)^{q/2} \right\}$$

and this leads to our result for some constant $c(q)$, because of the inequality (3.2). We can easily deduce this result by standard convexity inequalities, for $1 < q \leq 2$, from (3.2).

Let us introduce the sequence

$$\tau_{n,\gamma}^2 = C_\tau q s_{n,\gamma}^2 \left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}} \right) \log n,$$

where $\gamma < \beta$ are in B_{N_n} , $C_\tau > 0$ and $\tilde{\gamma}$ and $\tilde{\beta}$ are defined by (2.4).

Lemma 3.10 1. If the set B_{N_n} satisfies conditions (2.1) and (2.2), then

$$\frac{\log n}{\tilde{\beta}_N} \xrightarrow{n \rightarrow \infty} \infty, \log \tilde{\beta}_N \sqrt{\frac{\tilde{\beta}_N}{\log n}} \xrightarrow{n \rightarrow \infty} 0 \text{ and } \log \frac{1}{\Delta_n} \sqrt{\frac{\tilde{\beta}_N}{\log n}} \xrightarrow{n \rightarrow \infty} 0,$$

where $\tilde{\beta}_N = \tilde{\beta}_{N_n}$ is defined by the transformation (2.4).

2. If γ, β are in B_{N_n} such that $\gamma < \beta$ then there exist constants C_1, C_2 depending only on previously fixed constants such that

$$\begin{aligned} \sup_{f \in H_\beta} \frac{B_{n,\beta}^q(x_0, H_\beta) + s_{n,\beta}^q}{\psi_{n,\beta}^q} &\leq C_1, \\ \sup_{f \in H_\beta} \frac{B_{n,\gamma}^q(x_0, H_\beta) + \tau_{n,\gamma}^q}{\psi_{n,\beta}^q} &\leq C_2 \sqrt{\log n} \left(\frac{1}{n}\right)^{-\frac{1}{2}\left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}}\right)}. \end{aligned}$$

Proof. 1. The limits are easy consequences of hypotheses (2.1) and (2.2).

2. By Lemma 3.5, there exist b_{\max} and k_{\max} not depending on β , such that $b_\beta \leq b_{\max}$ and $\|K_\beta\|_2 \leq k_{\max}$, for any β in B_{N_n} . Thus, for $\beta \in B_-$ and $\gamma < \beta$:

$$\frac{B_{n,\beta}(x_0, H_\beta)}{\psi_{n,\beta}} \leq b_{\max}, \frac{s_{n,\beta}}{\psi_{n,\beta}} \leq k_{\max} \sqrt{\frac{\Lambda}{\log n}}$$

and

$$\frac{B_{n,\gamma}(x_0, H_\beta)}{\psi_{n,\beta}} \leq b_{\max} \left(\frac{\log n}{n}\right)^{-\frac{1}{2}\left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}}\right)}.$$

Finally,

$$\frac{\tau_{n,\gamma}}{\psi_{n,\beta}} \leq \frac{qC_\tau}{a\tilde{\beta}_1} \sqrt{\tilde{\beta}_N} \left(\frac{\log n}{n}\right)^{-\frac{1}{2}\left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}}\right)}.$$

Because $\tilde{\beta}_N / \log n \rightarrow 0$ when $n \rightarrow \infty$ we get the lemma for $\beta \in B_-$. For the case $\beta = \beta_{N_n}$, denoted β_N :

$$\frac{B_{n,\beta_N}(x_0, H_\beta)}{\psi_{n,\beta_N}} \leq b_{\max}, \frac{s_{n,\beta_N}}{\psi_{n,\beta_N}} \leq k_{\max} \sqrt{\Lambda}.$$

Moreover,

$$\begin{aligned} \frac{B_{n,\gamma}(x_0, H_\beta)}{\psi_{n,\beta_N}} &\leq b_{\max} \sqrt{\log n} \left(\frac{1}{n}\right)^{-\frac{1}{2}\left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}_N}\right)} \\ \frac{\tau_{n,\gamma}}{\psi_{n,\beta_N}} &\leq \frac{2qC_\tau}{a} \sqrt{\frac{\tilde{\beta}_{N_n}}{\log n}} \left(\frac{1}{n}\right)^{-\frac{1}{2}\left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\tilde{\beta}_N}\right)}. \end{aligned}$$

□

Lemma 3.11 *If f belongs to $H_\beta = H(\beta, p, L)$ and γ, β are elements of B_{N_n} such that $\gamma < \beta$, then*

$$\sup_{\substack{\gamma \in B_{N_n} \\ \gamma < \beta}} \frac{\beta_N}{\Delta_n} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [(Z_{n,\gamma}(x_0, H_\beta))^q I(Z_{n,\gamma} \geq \tau_{n,\gamma})] = o(1),$$

as $n \rightarrow \infty$, where $o(1)$ is independent of β in B_{N_n} .

Proof. Since $\|Z_{n,\gamma}(x_0, H_\beta)\|_\infty \leq K_{\max}/h_{n,\gamma}$ ($\|K_\gamma\|_\infty \leq K_{\max}$), we have

$$\begin{aligned} E_f [(Z_{n,\gamma}(x_0, H_\beta))^q I(Z_{n,\gamma} \geq \tau_{n,\gamma})] &= \int_{\tau_{n,\gamma}}^{K_{\max}/h_{n,\gamma}} P_f [Z_{n,\gamma} \geq u] du^q \\ &\leq \int_I + \int_{II}, \end{aligned}$$

where we need to split the integration domain in two intervals. In the rest of the proof we denote $c = c(q, \beta_1, L, \Lambda, K)$ a positive constant, that might depend on the other given constants. We give the proof only for $\beta \in B_-$. The case $\beta = \beta_N$ is similar.

Let $c_1 > 0$ be a constant, large enough such that $\tau_{n,\gamma} \leq c_1$ and let $I = [\tau_{n,\gamma}, c_1]$. We write

$$\begin{aligned} \int_I P_f [Z_{n,\gamma} \geq u] du^q &\leq 2 \int_{\tau_{n,\gamma}}^{c_1} \exp \left\{ -\frac{u^2}{2s_{n,\gamma}^2(1+c_0c_1)} \right\} du^q \\ &\leq (1+c_0c_1)^{q/2} s_{n,\gamma}^q \exp \left\{ -\frac{\tau_{n,\gamma}^2}{2s_{n,\gamma}^2(1+c_0c_1)} \right\}, \end{aligned}$$

where we applied Lemma 3.8. Then we use the first part of Lemma 3.10 to get

$$\begin{aligned} \frac{\beta_N}{\Delta_n} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} \int_I &\leq \frac{\beta_N}{\Delta_n} \frac{cs_{n,\gamma}^q}{\psi_{n,\beta}^q} \exp \left\{ -\frac{qC_\tau \log n}{2(1+c_0c_1)} \left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\beta} \right) \right\} \\ &\leq \frac{\beta_N}{\Delta_n} \frac{c}{(\log n)^{q/2}} \exp \left\{ -\frac{q \log n}{2} \frac{\Delta_n}{2\tilde{\beta}_{N_n}^2} \left(\frac{C_\tau}{(1+c_0c_1)} - 1 \right) \right\} \end{aligned}$$

which is on $o(1)$ if we choose $C_\tau > 0$ sufficiently large and does not depend on $\gamma < \beta$ in B_{N_n} .

At last, let us consider $c_2 \geq K_{\max}$ and $II = [c_1, c_2/h_{n,\gamma}]$. Then, by Lemma 3.8 and the fact that $\frac{u^2}{1+c_0u} \geq \frac{u}{2c_0}$ for u large enough:

$$\begin{aligned} \frac{\beta_N}{\Delta_n} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} \int_{II} &\leq \frac{\beta_N}{\Delta_n} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} \int_{c_1}^{\infty} 2 \exp \left\{ -\frac{u}{4c_0s_{n,\gamma}^2} \right\} du^q \\ &\leq \frac{2\beta_N}{\Delta_n} \sup_{f \in H_\beta} \left(\frac{4c_0s_{n,\gamma}^2}{\psi_{n,\beta}} \right)^q \exp \left\{ -\frac{c_1}{4c_0s_{n,\gamma}^2} \right\} \\ &\leq \frac{\beta_N}{\Delta_n} \left(\frac{\log n}{n} \right)^\alpha \frac{c}{(\log n)^q} \exp \left\{ -\frac{c_1 \log n}{4c_0c} \left(\frac{n}{\log n} \right)^{1-\frac{1}{2\beta_1}} \right\}, \end{aligned}$$

which is an $o(1)$ for some $\alpha > 0$ and the whole bound is free of γ and β over B_{N_n} . \square

Lemma 3.12 *Let f belong to the class $H_\beta = H(\beta, p, L)$ and define $\delta_{n1} = \exp\left\{-\frac{\Delta_n \log n}{8\beta_N^2}\right\}$. Then for arbitrary $\gamma_0, \gamma, \gamma_1$, elements of the set B_{N_n} such that $\gamma_0 \leq \gamma < \gamma_1 \leq \beta$ we have*

$$\lim_{n \rightarrow \infty} \frac{B_{n,\gamma_0}(x, H_\beta)}{\delta_{n1}\eta_{n,\gamma_0}} = 0 \text{ and } \lim_{n \rightarrow \infty} \frac{B_{n,\gamma_1}(x, H_\beta)}{\delta_{n1}\eta_{n,\gamma_0}} = 0.$$

Proof. We have that $\delta_{n1} \xrightarrow{n \rightarrow \infty} 0$ by **Assumption (A)**, relation (2.2). We also see that $h_{n,\gamma_0} \leq h_{n,\gamma_1}$, for n large enough. Then

$$\begin{aligned} \frac{B_{n,\gamma_0}(x, H_\beta)}{\delta_{n1}\eta_{n,\gamma_0}} &\leq \frac{B_{n,\gamma_1}(x, H_\beta)}{\delta_{n1}\eta_{n,\gamma_0}} \leq \frac{b_{n,\gamma_1} h_{n,\gamma_1}^{\tilde{\gamma}_1 - 1/2}}{\delta_{n1} a h_{n,\gamma_0}^{\tilde{\gamma}_0 - 1/2}} \leq \frac{C_3}{\delta_{n1}} \sqrt{\frac{h_{n,\gamma_0}}{h_{n,\gamma_1}}} \\ &\leq C_3 \exp\left\{\frac{\Delta_n \log n}{8\tilde{\beta}_N^2} - \log n \left(\frac{1}{4\tilde{\gamma}_0} - \frac{1}{4\tilde{\gamma}_1}\right)\right\} \leq C_3 \delta_{n1}^2, \end{aligned}$$

where $\tilde{\gamma}_0 = \tilde{\gamma}_0(H_\beta)$, $\tilde{\gamma}_1 = \tilde{\gamma}_1(H_\beta)$ and $\tilde{\beta}_N = \tilde{\beta}_N(H_\beta)$ are defined as in (2.4). This last term tends to 0 uniformly in γ_0, γ, β . □

Lemma 3.13 *If f belongs to $H_\beta = H(\beta, p, L)$ and $\gamma, \beta \in B_{N_n}$, $\gamma < \beta$, then there exist absolute positive constants C_4, C_η such that*

$$\sup_{f \in H_\beta} P_f \left[\hat{\beta} = \gamma \right] \leq C_4 \frac{\beta_N}{\Delta_n} n^{-\frac{q}{4\tilde{\gamma}} C_\eta}.$$

Proof. Let $\gamma_1 = \min\{\beta' \in B_{N_n} : \beta' > \gamma\}$, which implies that $\gamma_1 \leq \beta$. From the assumptions that $\hat{\beta} = \gamma$ and $\gamma < \beta$ we deduce that there must exist γ_0 in B_{N_n} such that $\gamma_0 < \gamma_1$ and $|f_{n,\gamma_1}(x_0) - f_{n,\gamma_0}(x_0)| > \eta_{n,\gamma_0}$. For brevity we do not mention the point x and the class H_β in the bias and stochastic term notation. Then

$$\begin{aligned} P_f \left[\hat{\beta} = \gamma \right] &\leq \sum_{\substack{\gamma_0 \in B_{N_n} \\ \gamma_0 \leq \gamma}} P_f \left[|f_{n,\gamma_1}(x_0) - f_{n,\gamma_0}(x_0)| > \eta_{n,\gamma_0} \right] \\ &\leq \text{card} B_{N_n} \sup_{\substack{\gamma_0 \in B_{N_n} \\ \gamma_0 < \gamma}} P_f \left[|f_{n,\gamma_1}(x_0) - f_{n,\gamma_0}(x_0)| > \eta_{n,\gamma_0} \right] \\ &\leq \frac{\beta_N}{\Delta_n} \sup_{\substack{\gamma_0 \in B_{N_n} \\ \gamma_0 < \gamma}} P_f \left[B_{n,\gamma_0} + Z_{n,\gamma_0} + B_{n,\gamma_1} + Z_{n,\gamma_1} > \eta_{n,\gamma_0} \right]. \end{aligned}$$

By Lemma 3.12, we can find a constant $d > 0$ such that $B_{n,\gamma_0} + B_{n,\gamma_1} \leq d\delta_{n1}\eta_{n,\gamma_0}$ and $\delta_{n1} \rightarrow 0$, for n large enough. Then we replace

$$\begin{aligned} P_f \left[\hat{\beta} = \gamma \right] &\leq \frac{\beta_N}{\Delta_n} \sup_{\substack{\gamma_0 \in B_{N_n} \\ \gamma_0 < \gamma}} P_f \left[Z_{n,\gamma_0} + Z_{n,\gamma_1} > \eta_{n,\gamma_0} (1 - d\delta_{n1}) \right] \\ &\leq \frac{\beta_N}{\Delta_n} \sup_{\substack{\gamma_0 \in B_{N_n} \\ \gamma_0 < \gamma}} \{ P_f \left[Z_{n,\gamma_0} > \eta_{n,\gamma_0} (1 - 2d\delta_{n1}) \right] + P_f \left[Z_{n,\gamma_1} > d\delta_{n1}\eta_{n,\gamma_0} \right] \}. \end{aligned}$$

We remark that we can write $\eta_{n,\gamma_0}^2 \geq qC_\eta s_{n,\gamma_0}^2 \log n / (2\tilde{\gamma}_0)$, where we can choose $C_\eta = a^2 / (q\Lambda K_{\max})$ sufficiently large. We apply twice Lemma 3.8 for $\tau_{n,\gamma_0} \leq \eta_{n,\gamma_0} (1 - 2d\delta_{n1})$

$$\begin{aligned} P_f [Z_{n,\gamma_0} > \eta_{n,\gamma_0} (1 - 2d\delta_{n1})] &\leq 2 \exp \left\{ -\frac{\eta_{n,\gamma_0}^2}{2s_{n,\gamma_0}^2} (1 - 2d\delta_{n1})^2 (1 - \delta_{n0}) \right\} \\ &\leq 2 \exp \left\{ -\frac{qC_\eta \log n}{4\tilde{\gamma}_0} (1 - d_1\delta_{n1}) \right\} \\ &\leq 2n^{-\frac{qC_\eta}{4\tilde{\gamma}}}. \end{aligned}$$

For the second right-hand side term, $d\delta_{n1}\eta_{n,\gamma_0} = o(1)$, then

$$\begin{aligned} P_f [Z_{n,\gamma_1} > d\delta_{n1}\eta_{n,\gamma_0}] &\leq 2 \exp \left\{ -\frac{d^2\delta_{n1}^2\eta_{n,\gamma_0}^2}{2s_{n,\gamma_1}^2} (1 - \delta_{n0}) \right\} \\ &\leq 2 \exp \left\{ -\frac{qC_\eta \log n}{4\tilde{\gamma}_0} \frac{d^2\delta_{n1}^2 s_{n,\gamma_0}^2}{s_{n,\gamma_1}^2} (1 - \delta_{n0}) \right\} \\ &= o\left(n^{-\frac{q}{4\tilde{\gamma}}C_\eta}\right), \end{aligned}$$

because

$$\begin{aligned} \frac{\delta_{n1}s_{n,\gamma_0}}{s_{n,\gamma_1}} &= \delta_{n1} \sqrt{\frac{h_{n,\gamma_1}}{h_{n,\gamma_0}}} \geq \delta_{n1} \left(\frac{n}{\log n}\right)^{\frac{\Delta_n}{4\beta_N^2}} \\ &\geq \exp \left\{ \frac{\Delta_n \log n}{4\tilde{\beta}_N^2} \left(1 - \frac{\log \log n}{\log n}\right) \right\} \xrightarrow{n \rightarrow \infty} \infty. \end{aligned}$$

Thus the lemma is proved. □

3.2 Proof of the upper bound

Let us start the proof of (2.8) by considering different possible cases for the estimator $\hat{\beta}(H_\beta)$ of β . We have

$$R_{n,\beta}(f_n^*, \psi_{n,\beta}, H_\beta) \leq R_{n,\beta}^+(f_n^*, \psi_{n,\beta}, H_\beta) + R_{n,\beta}^-(f_n^*, \psi_{n,\beta}, H_\beta), \quad (3.4)$$

where

$$\begin{aligned} R_{n,\beta}^+(f_n^*, \psi_{n,\beta}, H_\beta) &= \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f \left[|f_n^*(x_0) - f(x_0)|^q I(\hat{\beta} \geq \beta) \right] \\ R_{n,\beta}^-(f_n^*, \psi_{n,\beta}, H_\beta) &= \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f \left[|f_n^*(x_0) - f(x_0)|^q I(\hat{\beta} < \beta) \right]. \end{aligned}$$

Assume first that $\beta \in B_-$, the case when $\beta = \beta_{N_n}$ will be treated later. If we assume that $\hat{\beta} \geq \beta$, by the definition of $\hat{\beta}$ we have

$$|f_n^*(x_0) - f_{n,\beta}(x_0)| \leq \eta_{n,\beta}$$

and thus

$$\begin{aligned}
R_{n,\beta}^+(f_n^*, \psi_{n,\beta}, H_\beta) &\leq \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [(\eta_{n,\beta} + |f_{n,\beta}(x_0) - f(x_0)|)^q] \\
&\leq \sup_{f \in H_\beta} 2^{q-1} \left\{ \frac{\eta_{n,\beta}^q}{\psi_{n,\beta}^q} + 2^{q-1} \frac{B_{n,\beta}^q + E_f [Z_{n,\beta}]^q}{\psi_{n,\beta}^q} \right\} \\
&\leq 2^{q-1} \left\{ a^q + \sup_{f \in H_\beta} \frac{B_{n,\beta}^q + s_{n,\beta}^q}{\psi_{n,\beta}^q} \right\},
\end{aligned}$$

by Remark 3.9. Then, by Lemma 3.10 this bound is finite, uniformly in β over B_- , asymptotically:

$$\sup_{\beta \in B_-} R_{n,\beta}^+(f_n^*, \psi_{n,\beta}, H_\beta) \leq C_5.$$

By the inclusion of the events $\{\hat{\beta} < \beta\} \subseteq \bigcup_{\substack{\gamma \in B_{N_n} \\ \gamma < \beta}} \{\hat{\beta} = \gamma\}$, we may write:

$$\begin{aligned}
R_{n,\beta}^-(f_n^*, \psi_{n,\beta}, H_\beta) &\leq \sum_{\substack{\gamma \in B_- \\ \gamma < \beta}} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [|f_{n,\gamma}(x_0) - f(x_0)|^q I(\hat{\beta} = \gamma)] \\
&\leq \frac{\beta_{N_n}}{\Delta_n} \cdot \sup_{\substack{\gamma \in B_- \\ \gamma < \beta}} R_{n,\beta}^\gamma,
\end{aligned}$$

where $\text{card}B_{N_n} \leq \beta_{N_n} / \Delta_n$ and

$$R_{n,\beta}^\gamma = \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [|f_{n,\gamma}(x_0) - f(x_0)|^q I(\hat{\beta} = \gamma)].$$

We bound this risk as follows

$$\begin{aligned}
R_{n,\beta}^\gamma &\leq 2^{q-1} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} \left\{ B_{n,\gamma}^q \cdot P_f [\hat{\beta} = \gamma] + E_f [Z_{n,\gamma}^q \cdot I(\hat{\beta} = \gamma)] \right\} \\
&\leq 2^{q-1} \sup_{f \in H_\beta} \frac{B_{n,\gamma}^q}{\psi_{n,\beta}^q} \cdot \sup_{f \in H_\beta} P_f [\hat{\beta} = \gamma] + \\
&\quad + 2^{q-1} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [Z_{n,\gamma}^q \cdot I(\hat{\beta} = \gamma)].
\end{aligned}$$

Then

$$\begin{aligned}
&\sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [Z_{n,\gamma}^q \cdot I(\hat{\beta} = \gamma)] \\
&\leq 2^{q-1} \left\{ \sup_{f \in H_\beta} \frac{\tau_{n,\gamma}^q}{\psi_{n,\beta}^q} P_f [\hat{\beta} = \gamma] + \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f [Z_{n,\gamma}^q \cdot I(Z_{n,\gamma} \geq \tau_{n,\gamma})] \right\}.
\end{aligned}$$

We can see that

$$\begin{aligned} R_{n,\beta}^\gamma &\leq 2^{q-1} \sup_{f \in H_\beta} \frac{B_{n,\gamma}^q + \tau_{n,\gamma}^q}{\psi_{n,\beta}^q} \cdot \sup_{f \in H_\beta} P_f \left[\hat{\beta} = \gamma \right] + \\ &\quad + 2^{q-1} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f \left[Z_{n,\gamma}^q \cdot I(Z_{n,\gamma} \geq \tau_{n,\gamma}) \right]. \end{aligned} \quad (3.5)$$

We apply Lemmas 3.10 and 3.13 for the first term of the above right-hand side and get

$$\begin{aligned} &\frac{\beta_{N_n}}{\Delta_n} \sup_{\substack{\gamma \in B_- \\ \gamma < \beta}} \sup_{f \in H_\beta} \frac{B_{n,\gamma}^q + \tau_{n,\gamma}^q}{\psi_{n,\beta}^q} \cdot \sup_{f \in H_\beta} P_f \left[\hat{\beta} = \gamma \right] \\ &\leq C_5 \left(\frac{\beta_N}{\Delta_n} \right)^2 (\log n)^{q/2} \sup_{\substack{\gamma \in B_- \\ \gamma < \beta}} n^{\frac{q}{2} \left(\frac{1}{2\gamma} - \frac{1}{2\beta} \right)} n^{-\frac{q}{4\gamma} C_\eta} \\ &\leq C_5 \left(\frac{\beta_N}{\Delta_n} \right)^2 (\log n)^{q/2} n^{-\frac{q}{4\beta_N} C_\eta}, \end{aligned}$$

which tends to 0.

Finally, by Lemma 3.11

$$\frac{\beta_{N_n}}{\Delta_n} \sup_{\substack{\gamma \in B_- \\ \gamma < \beta}} \sup_{f \in H_\beta} \psi_{n,\beta}^{-q} E_f \left[(Z_{n,\gamma}(x_0, H_\beta))^q I(Z_{n,\gamma} \geq \tau_{n,\gamma}) \right] \xrightarrow{n \rightarrow \infty} 0,$$

uniformly in β over B_{N_n} . Then

$$\sup_{\beta \in B_-} R_{n,\beta}^-(f_n^*, \psi_{n,\beta}, H_\beta) = o(1).$$

As for the case where $\beta = \beta_{N_n}$ both denoted β_N , (3.4) still holds with

$$\begin{aligned} R_{n,\beta_N}^+(f_n^*, \psi_{n,\beta_N}, H_{\beta_N}) &= \sup_{f \in H_{\beta_N}} \psi_{n,\beta_N}^{-q} E_f \left[|f_n^*(x_0) - f(x_0)|^q I(\hat{\beta} = \beta_N) \right] \\ &= \sup_{f \in H_{\beta_N}} \psi_{n,\beta_N}^{-q} E_f \left[|f_{n,\beta_N}(x_0) - f(x_0)|^q \right] \\ &\leq \sup_{f \in H_{\beta_N}} \frac{B_{n,\beta_N}^q + c(q) s_{n,\beta_N}^q}{\psi_{n,\beta_N}^q} \end{aligned}$$

which is finite, by Lemma 3.10 and, respectively,

$$\begin{aligned} R_{n,\beta_N}^-(f_n^*, \psi_{n,\beta_N}, H_{\beta_N}) &= \sup_{f \in H_{\beta_N}} \psi_{n,\beta_N}^{-q} E_f \left[|f_n^*(x_0) - f(x_0)|^q I(\hat{\beta} < \beta_N) \right] \\ &\leq \frac{\beta_N}{\Delta_n} \cdot \sup_{\gamma \in B_-} R_{n,\beta_N}^\gamma = o(1). \end{aligned}$$

The proof goes similarly for the case $\beta \in B_-$. Indeed, we have the bound (3.5) for $\beta = \beta_N$ and we conclude by Lemmas 3.10, 3.13 and 3.11.

□

4 Lower bound

We shall prove that for any γ, β in the set B_{N_n} such that $\gamma < \beta$

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{\alpha \in \{\gamma, \beta\}} R_{n, \alpha} \left(\hat{f}_n, \psi_{n, \alpha}, H_\alpha \right) \geq c^q > 0, \quad (4.6)$$

where $H_\alpha = H(\alpha, p, L)$ denotes $W(\alpha, p, L)$ or $M(\alpha, p, L)$, respectively, and $c > 0$ is a constant. Suppose for the beginning that $\beta < \beta_{N_n}$.

Before we proceed to the proof of (4.6) let us construct functions f, K and f_h needed in our proof. Clearly, we can construct a compactly supported function $K \in H(\gamma, p, \frac{L}{2})$ with the support $[-D, D]$, such that $\|K\|_2 < \infty$, $\int_{\mathbb{R}} K(x) dx = 0$, $K(0) = c > 0$ and $\|K\|_\infty \leq K(0)$. Let us consider also f belonging to both $H(\gamma, p, L/2)$ and $H(\beta, p, L)$ such that f is positive on \mathbb{R} , and $f(x) = a$, in a neighborhood of x_0 , where $a > 0$ is fixed.

Define, for $\tilde{\gamma} = \tilde{\gamma}(H_\beta)$ defined in (2.4)

$$f_h(x) = f(x) + h^{\tilde{\gamma} - \frac{1}{2}} K\left(\frac{x - x_0}{h}\right),$$

where $h > 0$ will be specified later on. Then, $\int_{\mathbb{R}} f_h(x) dx = 1$ and $f_h > 0$ since h can be chosen such that $a > h^{\tilde{\gamma} - \frac{1}{2}} K(0)$. We can also see easily, that

$$\left\| h^{\tilde{\gamma} - \frac{1}{2}} K\left(\frac{\cdot - x_0}{h}\right) \right\|_2 \leq h^{\tilde{\gamma}} \|K\|_2.$$

Let us prove that f_h belongs to $H(\gamma, p, L)$. If $H_\gamma = W(\gamma, p, L)$,

$$\|f_h^{(\gamma)}\|_p \leq \|f^{(\gamma)}\|_p + \|K^{(\gamma)}\|_p \leq L,$$

while for $H_\gamma = M(\gamma, p, L)$

$$\begin{aligned} \int_{\mathbb{R}} |\mathcal{F}(f_h)(x)|^p |x|^{p\gamma} dx &\leq 2^{p-1} \int_{\mathbb{R}} |\mathcal{F}(f)(x)|^p |x|^{p\gamma} dx \\ &\quad + 2^{p-1} h^{p\gamma+1} \int_{\mathbb{R}} |\mathcal{F}(K)(hx)|^p |x|^{p\gamma} dx \\ &\leq L^p. \end{aligned}$$

Finally, let us choose

$$\xi = \frac{q}{2} \left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\beta} \right) \text{ and } h = h_{n, \gamma} = \left(\xi^2 \frac{\log n}{n} \right)^{\frac{1}{2\tilde{\gamma}}}.$$

Lemma 4.14 *Let $0 < \delta < 1$ be fixed. If X_1, \dots, X_n are i.i.d. observations of probability density f_h in $H(\gamma, p, L)$, as described above, and if $\tau = n^{-\xi(1-\delta_0)}$, for fixed $0 < \delta_0 < 1$, then one can choose δ_0 and a so that*

$$P_{f_h} \left[\prod_{i=1}^n \frac{f(X_i)}{f_h(X_i)} \geq \tau \right] \geq 1 - \delta.$$

Proof. We write

$$P_{f_h} \left[\prod_{i=1}^n \frac{f(X_i)}{f_h(X_i)} \geq \tau \right] = P_{f_h} \left[\sum_{i=1}^n \log \frac{f(X_i)}{f_h(X_i)} \geq \log \tau \right]. \quad (4.7)$$

Let us note that for any $0 < \delta < 1$ we can find $a_0(\delta) \in (0, 1)$ such that

$$-x - \frac{x^2}{2}(1 + \delta) \leq \log(1 - x) \leq -x,$$

for all $0 < x < a_0(\delta)$.

But $f_h(x_0) = f(x_0) + h^{\tilde{\gamma}-1/2}K(0) \geq a/2$ for n large enough. Then, for h small enough,

$$\begin{aligned} E_{f_h} \left[\log \frac{f(X_1)}{f_h(X_1)} \right] &= \int \log \left(1 - \frac{h^{\tilde{\gamma}-1/2}}{f_h(x)} K \left(\frac{x - x_0}{h} \right) \right) f_h(x) dx \\ &\geq \int -h^{\tilde{\gamma}-1/2} K \left(\frac{x - x_0}{h} \right) dx \\ &\quad - \frac{1}{2}(1 + \delta) \int \frac{h^{2\tilde{\gamma}-1}}{f_h(x)} K^2 \left(\frac{x - x_0}{h} \right) dx \\ &\geq -(1 + \delta) h^{2\tilde{\gamma}} \frac{1}{h} \int K^2 \left(\frac{x - x_0}{h} \right) \frac{dx}{a}. \end{aligned}$$

In conclusion,

$$E_{f_h} \left[\log \frac{f(X_1)}{f_h(X_1)} \right] \geq -(1 + \delta) \frac{\|K\|_2^2}{a} h^{2\tilde{\gamma}} \geq -(1 + \delta) \frac{\|K\|_2^2}{a} \xi^2 \frac{\log n}{n}, \quad (4.8)$$

for h small enough.

By similar considerations, we also get, for h small enough:

$$\begin{aligned} \sigma_1^2 &= \text{Var}_{f_h} \left[\log \frac{f(X_1)}{f_h(X_1)} \right] \leq E_{f_h} \left[\log^2 \frac{f(X_i)}{f_h(X_i)} \right] \\ &\leq \int \log^2 \left(1 - \frac{h^{\tilde{\gamma}-1/2}}{f_h(x)} K \left(\frac{x - x_0}{h} \right) \right) f_h(x) dx \\ &\leq \int (1 + \delta) \frac{h^{2\tilde{\gamma}-1}}{f_h(x)} K^2 \left(\frac{x - x_0}{h} \right) dx \\ &\leq (1 + \delta) \frac{2\|K\|_2^2}{a} h^{2\tilde{\gamma}} = (1 + \delta) \frac{2\|K\|_2^2}{a} \xi^2 \frac{\log n}{n}. \end{aligned} \quad (4.9)$$

Also

$$\begin{aligned} E_{f_h} \left[\log^2 \frac{f(X_i)}{f_h(X_i)} \right] &\geq (1 - \delta) \int \frac{h^{2\tilde{\gamma}-1}}{f_h(x)} K^2 \left(\frac{x - x_0}{h} \right) dx \\ &\geq (1 - \delta) \frac{\|K\|_2^2}{2a} h^{2\tilde{\gamma}} \\ &= (1 - \delta) \frac{\|K\|_2^2}{2a} \xi^2 \frac{\log n}{n}, \end{aligned} \quad (4.10)$$

because $f_h(x) \leq 2a$ for h small enough. Since $E_{f_h} [\log(f(X_1)/f_h(X_1))] \leq 0$ by Jensen's inequality then, by (4.8):

$$\left(E_{f_h} \left[\log \frac{f(X_i)}{f_h(X_i)} \right] \right)^2 = o \left(\frac{\|K\|_2^2 \xi^2 \log n}{a n} \right)$$

for n large enough and together with (4.10) this yields

$$\sigma_1^2 \geq (1 - \delta)^2 \frac{\|K\|_2^2 \xi^2 \log n}{2a n}, \quad (4.11)$$

for h small enough. Finally

$$\begin{aligned} E_{f_h} \left[\left| \log \frac{f(X_1)}{f_h(X_1)} \right|^3 \right] &\leq \int (1 + \delta) \left| \frac{h^{\tilde{\gamma}-1/2}}{f_h(x)} K \left(\frac{x-x_0}{h} \right) \right|^3 f_h(x) dx \\ &\leq (1 + \delta) h^{3\tilde{\gamma}-1/2} \frac{1}{h} \int \left| K \left(\frac{x-x_0}{h} \right) \right|^3 \frac{dx}{f_h^2(x)} \\ &\leq (1 + \delta) h^{3\tilde{\gamma}-1/2} \frac{4 \int |K|^3}{a^2} \\ &= (1 + \delta) \frac{4 \int |K|^3}{a^2} \left(\xi^2 \frac{\log n}{n} \right)^{\frac{3\tilde{\gamma}-1/2}{2\tilde{\gamma}}}, \end{aligned} \quad (4.12)$$

for h small enough.

Then, going back to (4.7)

$$\begin{aligned} &P_{f_h} \left[\prod_{i=1}^n \frac{f(X_i)}{f_h(X_i)} \geq \tau \right] \\ &= P_{f_h} \left[\frac{1}{\sigma_1 \sqrt{n}} \sum_{i=1}^n \left[\log \frac{f(X_i)}{f_h(X_i)} - E_{f_h} \left[\log \frac{f(X_i)}{f_h(X_i)} \right] \right] \geq m_n \right] \\ &= P_{f_h} [U_n \geq m_n], \end{aligned}$$

where

$$U_n = \frac{1}{\sigma_1 \sqrt{n}} \sum_{i=1}^n \left[\log \frac{f(X_i)}{f_h(X_i)} - E_{f_h} \left[\log \frac{f(X_i)}{f_h(X_i)} \right] \right]$$

is a sum of i.i.d. centered variables, $\text{Var}_{f_h} [U_n] = 1$ and

$$m_n = \frac{1}{\sigma_1 \sqrt{n}} \left(\log \tau - n E_{f_h} \left[\log \frac{f(X_1)}{f_h(X_1)} \right] \right).$$

Moreover, by (4.11) and (4.12)

$$\begin{aligned} &\frac{1}{(\sigma_1 \sqrt{n})^3} \sum_{i=1}^n E_{f_h} \left| \log \frac{f(X_i)}{f_h(X_i)} - E_{f_h} \left[\log \frac{f(X_i)}{f_h(X_i)} \right] \right|^3 \\ &\leq \frac{2^3 n}{\sigma_1^3 n^{3/2}} E_{f_h} \left| \log \frac{f(X_1)}{f_h(X_1)} \right|^3 \\ &\leq 2^3 \frac{(1 + \delta)}{(1 - \delta)^{3/2}} \frac{\sqrt{2} \int |K|^3}{\sqrt{a} \|K\|_2^3} \frac{1}{\xi \sqrt{\log n}} \left(\xi^2 \frac{\log n}{n} \right)^{\frac{\tilde{\gamma}-1/2}{2\tilde{\gamma}}} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

By Lyapounov's theorem, $U_n \xrightarrow{D} U$, where U is a standard Gaussian random variable.

We apply also (4.8) and (4.9) and choose $a > 0$ sufficiently large in our hypothesis so that:

$$\begin{aligned} m_n &= \frac{1}{\sigma_1 \sqrt{n}} \left[\log \tau - n E_{f_h} \left[\log \frac{f(X_1)}{f_h(X_1)} \right] \right] \\ &\leq \frac{-\xi \log n (1 - \delta_0) + (1 + \delta) \|K\|_2^2 \xi^2 \log n / a}{(1 - \delta) \xi \sqrt{\log n} \|K\|_2 / \sqrt{2a}} \\ &\leq -\frac{\sqrt{2a} \sqrt{\log n}}{(1 - \delta) \|K\|_2} \left(1 - \delta_0 - \frac{(1 + \delta)}{a} \|K\|_2^2 \xi \right) \xrightarrow{n \rightarrow \infty} -\infty. \end{aligned}$$

In conclusion, $P_{f_h} [U_n \geq m_n] \xrightarrow{n \rightarrow \infty} 1$, which proves the lemma. □

The left-hand side expression in the inequality (4.6) is:

$$\begin{aligned} LHS &= \liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \max \left\{ R_{n,\gamma} \left(\hat{f}_n, \psi_{n,\gamma}, H_\gamma \right), R_{n,\beta} \left(\hat{f}_n, \psi_{n,\beta}, H_\beta \right) \right\} \\ &\geq \liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \max \left\{ \psi_{n,\gamma}^{-q} E_{f_h} \left[\left| \hat{f}_n(x_0) - f_h(x_0) \right|^q \right], \right. \\ &\quad \left. \psi_{n,\beta}^{-q} E_f \left[\left| \hat{f}_n(x_0) - f(x_0) \right|^q \right] \right\} \\ &\geq \liminf_{n \rightarrow \infty} \inf_{T_n} \max \left\{ E_{f_h} [|T_n - \theta|^q], q_n^q E_f [|T_n|^q] \right\}, \end{aligned} \tag{4.13}$$

where

$$\begin{aligned} T_n &= \psi_{n,\gamma}^{-1} \left| \hat{f}_n(x_0) - f(x_0) \right|, \\ q_n &= \frac{\psi_{n,\gamma}}{\psi_{n,\beta}} \text{ and } \theta = \psi_{n,\gamma}^{-1} (f_h(x_0) - f(x_0)) = \frac{h^{\tilde{\gamma}-\frac{1}{2}}}{\psi_{n,\gamma}} K(0). \end{aligned}$$

Let us denote $R_n(T_n, \theta)$ the right-hand side expression in (4.13) and state the following lemma (Tsybakov [27]) that will help us to conclude.

Lemma 4.15 *Let $q_n, q > 0$, $\tau > 0$, $0 < \delta < \frac{1}{2}$ be real numbers and f, f_h be such that $|\theta| \geq c > 0$ and*

$$P_{f_h} \left[\prod_{i=1}^n \frac{f(X_i)}{f_h(X_i)} \geq \tau \right] \geq 1 - \delta. \tag{4.14}$$

Then

$$R_n(T_n, \theta) \geq \frac{(1 - \delta) (c - \delta)^q \tau q_n^q \delta^q}{\tau q_n^q \delta^q + (c - \delta)^q}.$$

Proof. We remark that $|\theta| \geq \xi^{1-1/(2\tilde{\gamma})} K(0) \geq c > 0$ since $\tilde{\gamma} > \frac{1}{2}$ for both classes of functions. Another useful remark is that $\{|T_n| \leq \delta\} \subseteq \{|T_n - \theta| \geq c - \delta\}$, for δ sufficiently small. Then

$$\begin{aligned} R_n(T_n, \theta) &\geq \inf_{T_n} \max \{(c - \delta)^q P_{f_h} [|T_n - \theta| \geq c - \delta], q_n^q \delta^q P_f [|T_n| \geq \delta]\} \\ &\geq \inf_{T_n} \max \{(c - \delta)^q P_{f_h} [|T_n| \leq \delta], q_n^q \delta^q P_f [|T_n| \geq \delta]\}. \end{aligned}$$

We write

$$\begin{aligned} P_f [|T_n| \geq \delta] &= E_{f_h} \left[I [|T_n| \geq \delta] \frac{dP_f}{dP_{f_h}} \right] \\ &\geq \tau \cdot E_{f_h} \left[I [|T_n| \geq \delta] \cdot I \left[\frac{dP_f}{dP_{f_h}} \leq \tau \right] \right] \\ &\geq \tau \cdot \left[P_{f_h} [|T_n| \geq \delta] - P_{f_h} \left[\frac{dP_f}{dP_{f_h}} \leq \tau \right] \right] \\ &\geq \tau \cdot [P_{f_h} [|T_n| \geq \delta] - \delta], \end{aligned}$$

where we applied the condition (4.14). Now

$$\begin{aligned} R_n(T_n, \theta) &\geq \inf_{T_n} \max \{(c - \delta)^q (1 - P_{f_h} [|T_n| \geq \delta]), \\ &\quad q_n^q \delta^q \tau (P_f [|T_n| \geq \delta] - \delta)\} \\ &\geq \inf_{0 \leq t \leq 1} \max \{(c - \delta)^q (1 - t), \tau q_n^q \delta^q (t - \delta)\} \\ &\geq \frac{(1 - \delta) (c - \delta)^q \tau q_n^q \delta^q}{\tau q_n^q \delta^q + (c - \delta)^q}. \end{aligned}$$

□

In order to finish the proof of (4.6), note that for τ introduced in Lemma 4.14:

$$\begin{aligned} \tau q_n^q &= n^{-\xi(1-\delta_0)} \left(\frac{\log n}{n} \right)^{q \left(\frac{1}{4\beta} - \frac{1}{4\tilde{\gamma}} \right)} \\ &= \exp \left\{ \log n \cdot \left[-\xi(1 - \delta_0) + \xi \left(1 - \frac{\log \log n}{\log n} \right) \right] \right\} \\ &= \exp \left\{ \frac{q \log n}{2} \left(\frac{1}{2\tilde{\gamma}} - \frac{1}{2\beta} \right) \left(\delta_0 - \frac{\log \log n}{\log n} \right) \right\} \end{aligned}$$

and that $\liminf_{n \rightarrow \infty} \tau q_n^q = \infty$ which implies, in view of Lemma 4.15 and of the arbitrariness of a small $\delta > 0$, that

$$\liminf_{n \rightarrow \infty} R_n(T_n, \theta) \geq c^q > 0.$$

In case $\beta = \beta_{N_n}$ (denoted β_N for brevity), the only change in the proof of the lower bound appears in the sequence

$$q_n = \frac{\psi_{n,\gamma}}{\psi_{n,\beta_N}} = (\log n)^{\frac{1}{2} - \frac{1}{4\beta_N}} \left(\frac{\log n}{n} \right)^{\left(\frac{1}{4\beta_N} - \frac{1}{4\tilde{\gamma}} \right)}$$

and we have also $\liminf_{n \rightarrow \infty} \tau q_n^q = \infty$ which brings us to the same conclusion.

□

5 Adaptive rate

Proof of Theorem 2.3. Recall that

$$\psi_{n,\beta} = \psi_{n,\beta}(H_\beta) = \begin{cases} (\log n/n)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}, & \text{if } \beta \in B_- \\ (1/n)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}, & \text{if } \beta = \beta_{N_n} \end{cases},$$

and that the optimal rate of convergence

$$\varphi_{n,\beta} = \varphi_{n,\beta}(H_\beta) = (1/n)^{\frac{\tilde{\beta}-1/2}{2\tilde{\beta}}}.$$

Suppose there exists an estimator $\tilde{f}_n(x_0)$ of $f(x_0)$ which is optimal rate adaptive over a set $B \subseteq B_{N_n}$ that contains at least two different values, where B_{N_n} satisfies **Assumption (A)**. This means that for a constant $C > 0$ we have

$$\begin{aligned} C &\geq \limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta}(\tilde{f}_n, \varphi_{n,\beta}, H_\beta) \\ &\geq \limsup_{n \rightarrow \infty} \sup_{\beta \in B} \left(\frac{\psi_{n,\beta}^q}{\varphi_{n,\beta}^q} R_{n,\beta}(\tilde{f}_n, \psi_{n,\beta}, H_\beta) \right) \end{aligned}$$

But

$$\frac{\psi_{n,\beta}^q}{\varphi_{n,\beta}^q} = (\log n)^{\frac{q}{2}\left(1-\frac{1}{2\tilde{\beta}}\right)} \geq (\log n)^{q/2},$$

then

$$C \geq \liminf_{n \rightarrow \infty} \left((\log n)^{q/2} \inf_{\hat{f}_n} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, \psi_{n,\beta}, H_\beta) \right).$$

This leads to a contradiction since the right-hand side term tends to ∞ . Indeed, we saw in Section 4 that

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, \psi_{n,\beta}, H_\beta) \geq c^q > 0.$$

□

Let us prove now that $\psi_{n,\beta}(H_\beta)$ is the adaptive rate of convergence over B_{N_n} in **Assumption (A)**, in the sense of Definition 2.1. Suppose that there exist another rate $\rho_{n,\beta}$ and an estimator f_n^{**} such that

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B_{N_n}} R_{n,\beta}(f_n^{**}, \rho_{n,\beta}, H_\beta) \leq C' < \infty. \quad (5.15)$$

Moreover, suppose that there exists β' in B_{N_n} such that $\rho_{n,\beta'}/\psi_{n,\beta'} \rightarrow 0$, when $n \rightarrow \infty$. We have to find another β'' in B_{N_n} such that

$$\frac{\rho_{n,\beta'}}{\psi_{n,\beta'}} \cdot \frac{\rho_{n,\beta''}}{\psi_{n,\beta''}} \xrightarrow{n \rightarrow \infty} \infty. \quad (5.16)$$

Remark that β' belongs necessarily to B_- . Indeed, if $\beta = \beta_N$, ψ_{n,β_N} coincides with the optimal rate of convergence φ_{n,β_N} which can not be improved over the class $H(\beta_N, p, L)$. Since $\varphi_{n,\beta}$ satisfies the lower bound:

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{f}_n} R_{n,\beta}(\widehat{f}_n, \varphi_{n,\beta}, H_\beta) \geq c' > 0, \quad (5.17)$$

$\rho_{n,\beta}$ can not be faster than $\varphi_{n,\beta}$. Indeed, as a consequence of (5.15),

$$\begin{aligned} C' &\geq \limsup_{n \rightarrow \infty} \left(\frac{\varphi_{n,\beta}}{\rho_{n,\beta}} \right)^q R_{n,\beta}(f_n^{**}, \varphi_{n,\beta}, H_\beta) \\ &\geq \left(\limsup_{n \rightarrow \infty} \frac{\varphi_{n,\beta}}{\rho_{n,\beta}} \right)^q \liminf_{n \rightarrow \infty} \inf_{\widehat{f}_n} R_{n,\beta}(\widehat{f}_n, \varphi_{n,\beta}, H_\beta). \end{aligned}$$

This implies in particular, for $\beta = \beta'$, that there exists some $c_1 > 0$ such that $\rho_{n,\beta'} \geq c_1 \varphi_{n,\beta'}$, for n large enough.

We put $\beta'' = \beta_N = \beta_{N_n}$, $0 < \delta < 1$ small number and $\varepsilon_n = (1/n)^{1/2-\delta/\beta''}$. Two cases can possibly occur. At first, if

$$\frac{\rho_{n,\beta_N}}{\varepsilon_n} \xrightarrow[n \rightarrow \infty]{} \infty,$$

we have, for n large enough,

$$\begin{aligned} \frac{\rho_{n,\beta'}}{\psi_{n,\beta'}} \cdot \frac{\rho_{n,\beta_N}}{\psi_{n,\beta_N}} &\geq \frac{c_1 \varphi_{n,\beta'}}{\psi_{n,\beta'}} \cdot \frac{\rho_{n,\beta_N}}{\varepsilon_n} \cdot \frac{\varepsilon_n}{\psi_{n,\beta_N}} \\ &\geq c_1 \frac{\rho_{n,\beta_N}}{\varepsilon_n} \cdot \frac{n^{\frac{\delta}{\beta''} - \frac{1}{4\beta_N}}}{(\log n)^{1 - \frac{1}{4\beta''}}} \xrightarrow[n \rightarrow \infty]{} \infty, \end{aligned}$$

and the result (5.16) follows. Secondly, if

$$\liminf_{n \rightarrow \infty} \frac{\rho_{n,\beta_N}}{\varepsilon_n} \leq c_2 < \infty,$$

we are lead to a contradiction with our assumptions and $\psi_{n,\beta}$ is proved to be adaptive rate of convergence. More precisely, if $\rho_{n,\beta_N} \leq c_2 \varepsilon_n$, for n large enough, let us denote $\varepsilon'_n = (1/n)^{1/2-2\delta/\beta''}$. By (5.15), we have:

$$\begin{aligned} C' &\geq \limsup_{n \rightarrow \infty} \max \left\{ R_{n,\beta'}(f_n^{**}, \rho_{n,\beta'}, H_{\beta'}), R_{n,\beta_N}(f_n^{**}, \rho_{n,\beta_N}, H_{\beta_N}) \right\} \\ &\geq \limsup_{n \rightarrow \infty} \max \left\{ \left(\frac{\psi_{n,\beta'}}{\rho_{n,\beta'}} \right)^q R_{n,\beta'}(f_n^{**}, \psi_{n,\beta'}, H_{\beta'}), \right. \\ &\quad \left. \left(\frac{\varepsilon'_n}{\rho_{n,\beta_N}} \right)^q R_{n,\beta_N}(f_n^{**}, \varepsilon'_n, H_{\beta_N}) \right\} \\ &\geq \limsup_{n \rightarrow \infty} \min \left\{ \left(\frac{\psi_{n,\beta'}}{\rho_{n,\beta'}} \right)^q, \left(\frac{\varepsilon_n}{\rho_{n,\beta_N}} n^{\frac{\delta}{\beta''}} \right)^q \right\}. \\ &\liminf_{n \rightarrow \infty} \inf_{\widehat{f}_n} \max \left\{ R_{n,\beta'}(\widehat{f}_n, \psi_{n,\beta'}, H_{\beta'}), R_{n,\beta_N}(\widehat{f}_n, \varepsilon'_n, H_{\beta_N}) \right\}. \end{aligned} \quad (5.18)$$

By our assumption, $\psi_{n,\beta'}/\rho_{n,\beta'} \rightarrow \infty$ and

$$\frac{\varepsilon_n}{\rho_{n,\beta_N}} n^{\frac{\delta}{\beta'}} \geq \frac{1}{c_2} \exp \left\{ \frac{\delta \log n}{\tilde{\beta}_N} \right\} \rightarrow \infty,$$

in view of **Assumption (A)**. It suffices to prove that

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \max \left\{ R_{n,\beta'} \left(\hat{f}_n, \psi_{n,\beta'}, H_{\beta'} \right), R_{n,\beta_N} \left(\hat{f}_n, \varepsilon'_n, H_{\beta_N} \right) \right\} \geq c'' > 0, \quad (5.19)$$

in order to get a contradiction in (5.18). We use for this purpose the proof of the lower bound in Section 4. We construct similarly, densities $f \in H(\beta_N, p, L)$ and $f_h \in H(\beta', p, L)$, such that

$$|\theta| = \left| \psi_{n,\beta'}^{-1} (f_h - f)(x_0) \right| \geq c_3 > 0$$

and

$$P_{f_h} \left[\prod_{i=1}^n \frac{f(X_i)}{f_h(X_i)} \geq \tau \right] \geq 1 - \delta,$$

for some $\tau, \delta > 0$. If $q_n = \psi_{n,\beta'}/\varepsilon'_n$, then by Lemma 4.15

$$\inf_{\hat{f}_n} \max \left\{ R_{n,\beta'} \left(\hat{f}_n, \psi_{n,\beta'}, H_{\beta'} \right), R_{n,\beta_N} \left(\hat{f}_n, \varepsilon'_n, H_{\beta_N} \right) \right\} \geq \frac{(1-\delta)(c_3-\delta)^q \tau q_n^q \delta^q}{\tau q_n^q \delta^q + (c_3-\delta)^q}.$$

We put $\tau = \left(\frac{1}{n}\right)^{q(1-16\delta)/(4\tilde{\beta}')}$, then

$$\tau q_n^q \geq n^{\frac{2q\delta}{\beta_N}} (\log n)^{\frac{q}{2}} \rightarrow \infty$$

and (5.19) is proved. □

References

- [1] Besov, O.V., Il'in V.P. and Nikolski S.M. *Integral representations of functions and embedding theorems*. J. Wiley, (1978) New York
- [2] Bretagnolle, J. and Huber, C. *Estimation des densités: risque minimax*. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete (1979) 47 pp. 119-137
- [3] Brown, L.D. and Low, M.G. *A constrained risk inequality with application to non-parametric functional estimation*. Ann. of Statist. (1996) 24 pp. 2524-2535
- [4] Butucea, C. *Estimation non paramétrique adaptative de la densité de probabilité; Vitesses de convergence, constante exacte et résultats numériques*. (1999) Ph. D. Thesis, Paris 6 University

- [5] Donoho, D.L. *Statistical estimation and optimal recovery*. Ann. Statist. (1994) 22 pp. 238-270
- [6] Donoho, D.L., Johnstone, I., Kerkyacharian, G., Picard, D. *Wavelet shrinkage: Asymptopia?* J. Roy. Statist. Soc., B, (1995) 57 pp. 301-369
- [7] Donoho, D.L., Johnstone, I., Kerkyacharian, G., Picard, D. *Density estimation by wavelet thresholding*. Ann. Statist. (1996) 24 pp. 508-539
- [8] Donoho, D.L. and Liu R.C. *Geometrizing rates of convergence, III*. Ann. Statist. (1991) 19 pp. 668-701
- [9] Donoho, D.L. and Low, M.G. *Renormalization exponents and optimal pointwise rates of convergence*. Ann. Statist. (1992) 20 pp. 944-970
- [10] Efromovich, S.Yu. *Nonparametric estimation of a density with unknown smoothness*. Theory Prob. Appl. (1985) 30 pp. 557-568
- [11] Farrell, R.H. *On best obtainable asymptotic rates of convergence in estimation of a density function at a point*. Ann. Math. Statist. (1972) 43 pp. 170-180
- [12] Goldenshluger, A. and Nemirovski, A. *On spatially adaptive estimation of nonparametric regression*. Math. Meth. of Statist. (1997) 6 pp. 135-170
- [13] Golubev, G.K. *Nonparametric estimation of smooth probability densities in L_2* . Probl. Inform. Transm. (1992) 28 pp. 44-54
- [14] Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A.B. *Wavelets, Approximations and Statistical Applications*. Lecture Notes in Statistics 129, (1998) Springer-Verlag, New York
- [15] Ibragimov, I.A. and Hasminskii, R.Z. *Statistical estimation: Asymptotic theory*. Springer-Verlag, (1981) New York
- [16] Ibragimov, I.A. and Hasminskii, R.Z. *Bounds for the risks of nonparametric regression estimates*. Theory Prob. Appl. (1982) 27 pp. 84-99
- [17] Juditsky, A. *Wavelet estimators: Adapting to unknown smoothness*. Math. Meth. of Statist. (1997) 6 pp. 1-25
- [18] Kerkyacharian, G., Picard, D. and Tribouley, K. *L_p adaptive density estimation*. Bernoulli (1996) 2 pp. 229-247
- [19] Korostelev, A.P. and Tsybakov, A.B. *Minimax theory of image reconstruction*. Lecture Notes in Statistics 82, (1993) Springer-Verlag, New York
- [20] Lepskii, O.V. *On a problem of adaptive estimation in Gaussian white noise*. Theory Prob. Appl. (1990) 35 pp. 454-466
- [21] Lepskii, O.V. *Asymptotically minimax adaptive estimation. I: Upper bounds. Optimally adaptive estimates*. Theory Prob. Appl. (1991) 36 pp. 682-697

- [22] Lepskii, O.V. *Asymptotically minimax adaptive estimation. II: Schemes without optimal adaptation: adaptive estimators*. Theory Prob. Appl. (1992) 37 pp. 433-448
- [23] Lepski, O.V., Mammen, E., Spokoiny, V.G. *Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors*. Ann. Statist. (1997) 25 pp. 929-947
- [24] Stein, E. and Weiss, G. *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, (1971) Princeton
- [25] Stone, C.J. *Optimal rates of convergence for nonparametric estimators*. Ann. Statist. (1980) 8 pp. 1348-1360
- [26] Stone, C.J. *Optimal global rates of convergence for nonparametric regression*. Ann. Statist. (1982) 10 pp. 1040-1053
- [27] Tsybakov, A.B. *Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes*. Ann. Statist. (1998) 6
- [28] Wahba, G. *Optimal convergence properties of variable knot, kernel and orthogonal series methods for density estimation*. Ann. of Statist. (1975) 3 pp. 15-29