Minimax exact constant in sup-norm for nonparametric regression with random design

KARINE BERTIN*

Abstract: We consider the nonparametric regression model with random design. We study the estimation of a regression function f in the uniform norm assuming that f belongs to a Hölder class. We determine the minimax exact constant and an asymptotically exact estimator. They depend on the minimum value of the design density.

Key Words: nonparametric regression, minimax risk, minimax exact constant, uniform norm.

AMS subject classification: 62G07, 62G20

1 Introduction

We study the problem of estimating a nonparametric regression function f on $[0,1]$ from observations

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n, \tag{1}$$

for $n > 1$ where the X_i are independent random variables in $[0,1]$ and the ξ_i are independent zero-mean Gaussian random variables with known variance σ^2 and independent of the X_i. We suppose that f belongs to the Hölder smoothness class $\Sigma(\beta, L)$ with β and L positive constants defined by:

$$\Sigma(\beta, L) = \left\{ f : |f^{(m)}(x) - f^{(m)}(y)| \le L|x-y|^\alpha, \quad x, y \in \mathbb{R} \right\}, \tag{2}$$

where $m = \lfloor \beta \rfloor$ is an integer such that $0 < \alpha \le 1$ and $\alpha = \beta - m$. Moreover, we suppose that f is bounded by a fixed constant $Q > 0$, so that f belongs to $\Sigma_Q(\beta, L)$ where

$$\Sigma_Q(\beta, L) = \Sigma(\beta, L) \cap \{f : \|f\|_\infty \le Q\},$$

and $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$. We suppose that the X_i have a density μ w.r.t. the Lebesgue measure, μ belongs to a Hölder class $\Sigma(l, C)$ with $0 < l \le 1$ and $C > 0$, and there exists $\mu_0 > 0$ such that $\min_{x\in[0,1]} \mu(x) = \mu_0$.

An estimator $\theta_n = \theta_n(x)$ of f is a measurable function with respect to the observations (1) and defined for $x \in [0,1]$. We define the maximal risk with sup-norm loss of an estimator θ_n by

$$R_n(\theta_n) = \sup_{f\in\Sigma_Q(\beta,L)} \mathbb{E}_f \left(w\left(\frac{\|\theta_n - f\|_\infty}{\psi_n} \right) \right),$$

*Laboratoire de Probabilité et Modèles Aléatoires, Université Pierre et Marie Curie, 175 rue de Chevaleret, F-75013 Paris, France. Tel 0033-(0)1-44-27-85-10. Email: kbertin@ccr.jussieu.fr

where $w(u)$ is a continuous non-decreasing function defined for $u \geq 0$ which has a polynomial upper bound $w(u) \leq W_0 (1 + u^\gamma)$ with some positive constants W_0, γ and such that $w(0) = 0$, $\mathbb{E}_f$ is the expectation with respect to the joint distribution $\mathbb{P}_f$ of the (X_i, Y_i) and $\psi_n = \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$. Let us recall that in our model, ψ_n is the minimax rate of convergence (c.f. Ibragimov and Hasminskii (1981), Ibragimov and Hasminskii (1982) , Stone (1982)).

Our goal is to determine the minimax exact constant C and an estimator θ_n^* such that

$$w(C) = \lim_{n\to\infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n\to\infty} R_n(\theta_n^*), \tag{3}$$

where $\inf_{\theta_n}$ stands for the infimum over all the estimators. An estimator that satisfies (3) is said to be asymptotically exact. The aim of this paper is to extend a result proved by Korostelev (1993) to the regression model with random design. Korostelev (1993) studied the estimation of a function $f \in \Sigma(\beta, L)$ with $0 < \beta \leq 1$ with sup-norm loss and for the regression model with fixed equidistant design ($X_i = i/n$ in (1)). He obtained the exact constant which is $w(C_0)$ with

$$C_0 = \left(\sigma^{2\beta} L \left(\frac{\beta+1}{2\beta^2}\right)^\beta\right)^{\frac{1}{2\beta+1}}$$

and an asymptotically exact estimator which is a kernel estimator close to

$$\widehat{f}_n(t) = \frac{1}{nh} \sum_{i=1}^{n} Y_i K\left(\frac{t - i/n}{h}\right). \tag{4}$$

Here h is a bandwidth depending on n and

$$K(t) = \frac{\beta+1}{2\beta}(1 - |t|^\beta)_+ \tag{5}$$

with $x_+ = \max(0, x)$. Donoho (1994) extended Korostelev's result to the Gaussian white noise model for Hölder classes with $\beta > 1$ for estimation in sup-norm. He proved that asymptotically exact estimators and exact constants in several settings with Hölder classes (2) are closely related to the solution f_β of the optimization problem

$$\max_{\substack{\|f\|_2 \leq 1 \\ f \in \Sigma(\beta,1)}} f(0), \tag{6}$$

which is linked to an "optimal recovery problem". He proved that the asymptotically exact estimators are kernel estimators (the equivalent of (4) for white noise model) where the kernel K is expressed in terms of the solution f_β:

$$K(t) = \frac{f_\beta(t)}{\int f_\beta(s) ds}, \tag{7}$$

and that the exact constant depends on $f_\beta(0)$. For $0 < \beta \leq 1$, the solution of (6) is known (see Korostelev (1993) or Donoho (1994)) and the kernel used by Korostelev defined in (5) is equal to that defined in (7) up to a renormalization on the support. However the function f_β is not known for $\beta > 1$, except for $\beta = 2$. Korostelev and Nussbaum (1999) have found the exact constant and asymptotically exact estimator for the density model in sup-norm. Lepski (1992) has studied the exact constant in the case of adaptation for the white noise model. The sup-norm estimation is only one of the approaches studied in the nonparametric literature. For the L_2-norm risk, one can find overview of results on exact minimax and adaptive estimation in the books of Efromovich (1999) and Tsybakov (2004)

Our results are the following. In Section 2, we give an asymptotically exact estimator θ_n^* and the exact constant for the regression model with random design. If the density μ is uniform ($\mu_0 = 1$), then the constant is equal to $w(C_0)$ (the constant of Korostelev (1993)). As it could be expected, the exact constant and the asymptotically exact estimator θ_n^* depend on the minimum value of the design density μ_0. It means that the asymptotically minimax estimators contribute to the sup-norm risk essentially at the points where we have less observations. The estimator θ_n^* that is proposed in Section 2 is close to a Nadaraya-Watson estimator and is independent of Q. The proofs are given in Section 3.

2 The main result and the estimator

In this section, we define an estimator θ_n^*. We shall prove in Subsection 3.1 that θ_n^* is an asymptotically exact estimator. This estimator is close to a Nadaraya-Watson estimator with the kernel K defined in (5). The bandwidth of θ_n^* is

$$h = \left(\frac{C_0'\psi_n}{L}\right)^{\frac{1}{\beta}},$$

with

$$C_0' = \left(\sigma^{2\beta} L \left(\frac{\beta+1}{2\beta^2\mu_0}\right)^{\beta}\right)^{\frac{1}{2\beta+1}}.$$

First let us define θ_n^* in a regular grid of points $x_k = \frac{km}{n} \in [0,1]$ for $k \in \{1, \dots, [\frac{n}{m}]\}$, with $m = [\delta_n n \psi_n^{\frac{1}{\beta}} + 1]$, $\delta_n = \frac{1}{\log n}$ and $[x]$ denotes the integer part of x. To account for the boundary effects, we need to introduce other kernels:

$$K_1(t) = 2K(t)I_{[0,1]}(t), \quad K_2(t) = 2K(t)I_{[-1,0]}(t) \quad \text{for} \quad t \in \mathbb{R}.$$

The estimator θ_n^* is defined for $k \in \{1, \dots, [\frac{n}{m}]\}$ by

$$\theta_n^*(x_k) = \frac{\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) Y_j}{\max\left(\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right), \delta_n\right)}, \tag{8}$$

if $x_k \in [h, 1-h]$. If $x_k \in [0,h)$ (respectively $x_k \in (1-h, 1]$), $\theta_n^*(x_k)$ is defined by (8) where K is replaced by K_1 (respectively by K_2). Finally the function θ_n^* is defined to be the polygonal function connecting the points $(x_k, \theta_n^*(x_k))$ for $k \in \{1, \dots, [\frac{n}{m}]\}$. Moreover, we put $\theta_n^*(x) = \theta_n^*(x_1)$ if $x \in [0, x_1]$ and if $x_{[\frac{n}{m}]} < 1$ we put $\theta_n^*(x) = \theta_n^*(x_{[\frac{n}{m}]})$ for $x \in [x_{[\frac{n}{m}]}, 1]$.
The results we obtain are the following:

Theorem 1. *We consider the model and the assumptions defined in Section 1. We suppose that the function $f \in \Sigma_Q(\beta, L)$, with $0 < \beta \leq 1$. The estimator θ_n^* satisfies*

$$\lim_{n\to\infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n\to\infty} R_n(\theta_n^*) = w(C_0').$$

We are going to prove Theorem 1 in two steps: the upper bound (Subsection 3.1) and the lower bound (Subsection 3.2). Let $0 < \varepsilon < 1/2$. In Subsection 3.1, we show that θ_n^* satisfies

$$\limsup_{n\to\infty} \sup_{f\in\Sigma_Q(\beta,L)} \mathbb{E}_f \left[w \left(\|\theta_n^* - f\|_\infty \psi_n^{-1} \right) \right] \leq w \left(C_0' \left(1+\varepsilon\right)^2 \right). \tag{9}$$

In Subsection 3.2, we prove that

$$\liminf_{n\to\infty} \inf_{\theta_n} \sup_{f\in\Sigma_Q(\beta,L)} \mathbb{E}_f [w(\|\theta_n - f\|_\infty \psi_n^{-1})] \geq w(C_0'(1-\varepsilon)). \tag{10}$$

Since $0 < \varepsilon < 1/2$ in (9) and (10) can be arbitrarily small and w is a continuous function, this proves Theorem 1.

Remarks:

(i) We introduce the cut-off δ_n in (8) to account for the case $\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) = 0$ which leads to a zero denominator. As it is proved in Lemma 1, $\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) - \mu(x_k)$ tends to 0 in probability as soon as n tends to ∞, so that $\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) = 0$ essentially does not occur.

(ii) The estimator θ_n^* does not depend on Q, but it depends on μ_0. It is possible to construct an asymptotically exact estimator independent of μ_0 and Q but the proof is rather technical. For this purpose, we cut the sample $(X_1, \ldots, X_n)$ in two parts of size α_n and $n - \alpha_n$, where α_n is an integer such that $\alpha_n \to \infty$ and $\alpha_n / n \to 0$ as $n \to \infty$. We estimate μ_0 with the part $(X_1, \ldots, X_{\alpha_n})$ of the sample by

$$\hat{\mu}_0 = \min_{k=1,\ldots,n} \hat{\mu}_n \left(\frac{k}{n} \right),$$

where $\hat{\mu}_n(x) = \frac{1}{\alpha_n g_n} \sum_{i=1}^{\alpha_n} K\left(\frac{X_i - x}{g_n}\right)$ and g_n such that $g_n \to 0$ and $\alpha_n g_n \to \infty$. We construct an estimator of f in the same way as θ_n^* except we only use the part $(X_{\alpha_n+1}, \ldots, X_n)$ of the sample and we replace μ_0 by $\hat{\mu}_0$, if the latter is not zero, in C_0' and h. The results are similar to those of this paper for this estimator but one needs to consider the law conditioned by $(X_1, \ldots, X_{\alpha_n})$.

(iii) We have only solved the problem of exact constant and asymptotically exact estimator for the Hölder classes $\Sigma_Q(\beta, L)$ such that $0 < \beta \leq 1$. In this case we have an explicit form for the constant and the estimator. An extension to $\beta > 1$ is possible but it does not give realizable estimators (since the solution f_β of the problem (6) is not explicitly known except for $\beta = 2$). A similar result could be found and the exact constant will be C_1 with

$$C_1 = f_\beta(0) \left(\sigma^{2\beta} L \left(\frac{2}{\mu_0(2\beta+1)} \right)^\beta \right)^{\frac{1}{2\beta+1}}.$$

The analogue of inequality (9) for $\beta > 1$ holds for example for the estimator θ_n^* defined for $t \in [0,1]$ by:

$$\theta_n^*(t) = \frac{\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - t}{h}\right) Y_j}{\max\left(\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - t}{h}\right), \delta_n\right)}, \tag{11}$$

with certain modifications near the boundaries. To prove inequality (9) with this new estimator, we will use methods similar to those of Lepski and Tsybakov (2000), based on the supremum of Gaussian processes. For $\beta > 1$, the proof of inequality (10) is the same as that of Subsection 3.2, but we need to use the function f_β and the fact that f_β is compactly supported. This was proved by Leonov (1997). He also proved that f_β is continuous and even for all $\beta > 1$.
(iv) Our result can be presumably extended to the white noise model

$$dY(t) = f(t)dt + \sigma(t)dW(t), \quad t \in [0,1],$$

where W is a standard Wiener process and σ^{-2} serves to replace the density of the design points. In this model, the maximum of σ^2 corresponds to the minimum value of the design density μ_0. An asymptotically exact estimator will be of the form

$$\theta_n^*(t) = \frac{1}{h}\int K\left(\frac{u-t}{h}\right) dY(u),$$

where h a bandwidth that depends on n and K is defined in (7).
(v) The constants L and β are supposed to be known, but using the techniques similar to Lepski (1992), one can presumably obtain adaptive asymptotically exact estimator. One should note however that the exact constant for adaptive estimator would be in general different.

3 Proofs

3.1 Proof of inequality (9)

We define the event A_n as

$$A_n = \left\{ \max_{x_k \in [h,1-h]} \left| \mu(x_k) - \frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) \right| < \delta_n \right\}.$$

Similarly we define $A_{1,n}$ (respectively $A_{2,n}$) which are obtained by replacing K by K_1 (respectively by K_2) and taking the supremum over $x_k \in [0,h)$ (respectively over $x_k \in (1-h,1]$). We define also

$$A'_n = \left\{ \max_{x_k \in [h,1-h]} \left| \frac{\mu(x_k)(\beta+1)}{2\beta+1} - \frac{1}{nh}\sum_{j=1}^n K^2\left(\frac{X_j - x_k}{h}\right) \right| < \delta_n \right\},$$

and the events $A'_{1,n}$ (respectively $A'_{2,n}$) obtained by replacing K by K_1 (respectively by K_2) and taking the supremum over $x_k \in [0,h)$ (respectively over $x_k \in (1-h,1]$). Let $B_n = A_n \cap A_{1,n} \cap A_{2,n} \cap A'_n \cap A'_{1,n} \cap A'_{2,n}$. We have the following result.

Lemma 1. *There exists $c > 0$ such that*

$$\mathbb{P}^X(B_n) \geq 1 - 12\frac{n}{m}\exp\left(-cnh\delta_n^2\right),$$

for n large enough, where P^X is the joint distribution of $X = (X_1, \dots, X_n)$.

The proof of the lemma is given in Subsection 3.3.

Before proving inequality (9), we give four propositions studying the behaviour of $\Delta_n = \psi_n^{-1}\|f-\theta_n^*\|_\infty$ on B_n and B_n^C. We postpone their proofs to Subsection 3.3. We denote I_B the indicator function of a set B which takes the value 1 on B and 0 otherwise. In the sequel, D_0, D_1,... are positive constants.

Proposition 1. *We have*

$$\lim_{n\to\infty} \sup_{f\in\Sigma_Q(\beta,L)} \mathbb{E}_f\left[w\left(\psi_n^{-1}\|f-\theta_n^*\|_\infty\right) I_{B_n^C}\right] = 0.$$

Define the bias and the stochastic terms for $x \in [0,1]$

$$b_n(x,f) = \mathbb{E}_f(\theta_n^*(x)I_{B_n}) - f(x)P^X(B_n),$$
$$Z_n(x,f) = \theta_n^*(x)P^X(B_n) - \mathbb{E}_f(\theta_n^*(x)I_{B_n}).$$

Proposition 2. *The bias term satisfies*

$$\limsup_{n\to\infty} \sup_{f\in\Sigma_Q(\beta,L)} \psi_n^{-1}\|b_n(.,f)\|_\infty \le \frac{C_0'}{2\beta+1}.$$

We study the stochastic term at the points x_k and we take n large enough such that $P^X(B_n) > 0$. For $k \in \{1,\ldots,[\frac{n}{m}]\}$, the process $Z_n(\cdot,f)$ satisfies $Z_n(x_k,f)I_{B_n} = \left(\widetilde{Z}_n(x_k,f)+\widehat{Z}_n(x_k,f)\right)I_{B_n}$, where for $x_k\in[h,1-h]$, on B_n

$$\widehat{Z}_n(x_k,f) = \frac{\frac{1}{nh}\sum_{j=1}^n \xi_j K\left(\frac{X_j-x_k}{h}\right)}{\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j-x_k}{h}\right)} P^X(B_n),$$

$$\widetilde{Z}_n(x_k,f) = U_n(x_k,f)P^X(B_n) - \mathbb{E}_f(U_n(x_k,f)I_{B_n}),$$

with

$$U_n(x_k,f) = \frac{\frac{1}{nh}\sum_{j=1}^n f(X_j) K\left(\frac{X_j-x_k}{h}\right)}{\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j-x_k}{h}\right)}.$$

For $x_k\in[0,h)$ (respectively $(1-h,1]$), $\widehat{Z}_n(x_k,f)$, $\widetilde{Z}_n(x_k,f)$ and $U_n(x_k,f)$ are defined in the same way except that we replace K by K_1 (respectively K_2).

Proposition 3. *The process $\widehat{Z}_n(\cdot,f)$ satisfies for all $z>1$ and n large enough*

$$\sup_{f\in\Sigma_Q(\beta,L)} \mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k\left|\widehat{Z}_n(x_k,f)\right| > \frac{2\beta C_0' z}{2\beta+1}\right\}\cap B_n\right] \le \delta_n^{-1}(\log n)^{-\frac{1}{2\beta+1}} n^{-\alpha_1(n)},$$

where $\alpha_1(n) = \frac{z^2C(\delta_n)-1}{2\beta+1}$ and $C(\delta_n)$ tends to 1 as $n\to\infty$.

Proposition 4. *The process $\widetilde{Z}_n(\cdot,f)$ satisfies for $z\ge\varepsilon/2$*

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k\left|\widetilde{Z}_n(x_k,f)\right| > \frac{2\beta C_0' z}{2\beta+1}\right\}\cap B_n\right] \le 2\delta_n^{-1}\psi_n^{-1/\beta}\exp\left(-D_0 z\psi_n\right),$$

where D_0 is independent of $f\in\Sigma_Q(\beta,L)$.

Here we prove inequality (9). By Proposition 1, $\limsup_{n\to\infty} \mathbb{E}_f(w(\Delta_n)I_{B_n^C}) = 0$. We have, using the monotonicity of w

$$\mathbb{E}_f(w(\Delta_n)I_{B_n}) \le w(C_0'(1+\varepsilon)^2)\mathbb{P}_f\left[\Delta_n I_{B_n} \le C_0'(1+\varepsilon)^2\right]$$

$$+\left(\mathbb{E}_f\left(w^2(\Delta_n)I_{B_n}\right)\right)^{\frac{1}{2}}\left(\mathbb{P}_f\left[\Delta_n I_{B_n} > C_0'(1+\varepsilon)^2\right]\right)^{\frac{1}{2}}.$$

To obtain the inequality (9), it is enough to prove that
(i) $\lim_{n\to\infty}\sup_{f\in\Sigma_Q(\beta,L)}\mathbb{P}_f\left[\Delta_n I_{B_n} > C_0'(1+\varepsilon)^2\right] = 0$,
(ii) there exists a constant D_1 such that $\limsup_{n\to\infty}\sup_{f\in\Sigma_Q(\beta,L)}\mathbb{E}_f(w^2(\Delta_n)I_{B_n}) \le D_1$.

Here we prove (i). Considering n large enough such that $P^X(B_n) \ge \frac{1}{1+\varepsilon}$, we have

$$\mathbb{P}_f\left[\Delta_n I_{B_n} > C_0'(1+\varepsilon)^2\right] = \mathbb{P}_f\left[\Delta_n I_{B_n} P^X(B_n) > C_0' P^X(B_n)(1+\varepsilon)^2\right]$$

$$\le \mathbb{P}_f\left[\Delta_n I_{B_n} P^X(B_n) > C_0'(1+\varepsilon)\right].$$

Note also that

$$\Delta_n I_{B_n} P^X(B_n) \le \psi_n^{-1}(\|b_n(\cdot,f)\|_\infty + \|Z_n(\cdot,f)\|_\infty)I_{B_n}.$$

Thus using Proposition 2, we deduce that, for n large enough

$$\mathbb{P}_f\left[\Delta_n I_{B_n} > C_0'(1+\varepsilon)^2\right] \le \mathbb{P}_f\left[\left\{\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty > \frac{2\beta C_0'(1+\varepsilon)}{(2\beta+1)}\right\}\cap B_n\right].$$

Since θ_n^* is the polygonal function connecting the points $(x_k, \theta_n^*(x_k))$, $Z_n(\cdot,f)$ is the polygonal function connecting the points $(x_k, Z_n(x_k,f))$. Thus we have $\|Z_n(\cdot,f)\|_\infty = \max_k |Z_n(x_k,f)|$,

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty > \frac{2\beta C_0'(1+\varepsilon)}{2\beta+1}\right\}\cap B_n\right] = \mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k|Z_n(x_k,f)| > \frac{2\beta C_0'(1+\varepsilon)}{2\beta+1}\right\}\cap B_n\right],$$

and

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k|Z_n(x_k,f)| > \frac{2\beta C_0'(1+\varepsilon)}{2\beta+1}\right\}\cap B_n\right] \le \mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k|\widehat{Z}_n(x_k,f)| > \frac{2\beta C_0'(1+\varepsilon/2)}{2\beta+1}\right\}\cap B_n\right]$$

$$+\mathbb{P}_f\left[\left\{\psi_n^{-1}\max_k|\widetilde{Z}_n(x_k,f)| > \frac{2\beta C_0'\varepsilon/2}{2\beta+1}\right\}\cap B_n\right].$$

Since $C(\delta_n)$ tends to 1 as $n\to\infty$, in view of Propositions 3 and 4 used respectively with $z = 1+\varepsilon/2$ and $z = \varepsilon/2$, the right hand side of the last inequality tends to 0 as n tends to ∞ uniformly in $f \in \Sigma_Q(\beta,L)$. So we obtain (i).

Here we prove (ii). We have, since $w(u) \le W_0(1+u^\gamma)$,

$$\mathbb{E}_f\left(w^2(\Delta_n)I_{B_n}\right) \le D_2 + D_3\left[\mathbb{E}_f\left(\left(\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty\right)^{2\gamma} I_{B_n}\right) + \left(\psi_n^{-1}\|b_n(.,f)\|_\infty\right)^{2\gamma}\right](1+o(1)).$$

Using the fact that

$$\mathbb{E}_f\left(\left(\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty\right)^{2\gamma} I_{B_n}\right) = \int_0^{+\infty}\mathbb{P}_f\left[\left(\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty\right)^{2\gamma} I_{B_n} > t\right]dt,$$

Propositions 3 and 4, and noting that $C(\delta_n)$ tends to 1 as $n\to\infty$, we prove that $\limsup_{n\to\infty}\mathbb{E}_f\left[\left(\psi_n^{-1}\|Z_n(\cdot,f)\|_\infty\right)^{2\gamma}\right] < \infty$. This and Proposition 2 entail (ii).

3.2 Proof of inequality (10)

3.2.1 Preliminaries

First, we need to define Σ', a subspace of $\Sigma_Q(\beta, L)$. Since μ satisfies a Lipschitz condition on $[0,1]$, there exists $x_0 \in [0,1]$ such that $\mu(x_0) = \min_{x\in[0,1]} \mu(x)$. Let $\gamma_n = (n/\log n)^{-\frac{\varepsilon}{2\beta+1}}$ for ε introduced in Section 2. Let $M = \left[\frac{\gamma_n}{2h(2^{1/\beta}+1)}\right]$ and define the points $a_1, \ldots, a_M$ in a neighbourhood of x_0 in the following way. For n large enough and if $x_0 \in (0,1)$, we put

$$a_1 = x_0 - \gamma_n/2 + \left(2^{1/\beta}+1\right) h, \qquad a_{j+1} - a_j = 2\left(2^{1/\beta}+1\right) h.$$

If $x_0 = 0$ (respectively $x_0 = 1$), we define the points a_j in the same way except that a_1 is $\left(2^{1/\beta}+1\right) h$ (respectively $1 - \gamma_n + \left(2^{1/\beta}+1\right) h$). We define the set Σ' as

$$\Sigma' = \left\{ f(\cdot,\theta), \theta \in [-1,1]^M \right\}, \tag{12}$$

where for $\theta = (\theta_1, \ldots, \theta_M) \in [-1,1]^M$ and $x \in [0,1]$

$$f(x,\theta) = Lh^\beta \sum_{j=1}^{M} \theta_j \left(1 - \left|\frac{x-a_j}{h}\right|^\beta\right)_+ .$$

For all $\theta \in [-1,1]^M$, $f(\cdot,\theta) \in \Sigma(\beta, L)$(cf. the appendix) and $\|f\|_\infty \le Q$ for n large enough. Therefore for n large enough $\Sigma' \subset \Sigma_Q(\beta, L)$.
Remark: For $\beta > 1$ the subspace Σ' should be defined in a similar way:

$$\Sigma' = \left\{ f(x,\theta) = Lh^\beta \sum_{j=1}^{M} \theta_j f_\beta\left(\frac{x-a_j}{h}\right), \quad \theta \in [-1,1]^M \right\}$$

and the values $(a_j)_{j=1,\ldots,M}$ should satisfy $a_{j+1} - a_j = 2A_\beta h(2^{1/\beta}+1)$ where $[-A_\beta, A_\beta]$ is the support of f_β.

Then we need to introduce an event N_n that satisfies the following lemma

Lemma 2. *The event*

$$N_n = \left\{ (X_1, \ldots, X_n) : \sup_{j=1,\ldots,M} \left| \frac{(\beta+1)(2\beta+1)}{4\mu_0\beta^2 nh} \sum_{k=1}^{n} \left(1 - \left|\frac{X_k - a_j}{h}\right|^\beta\right)_+^2 - 1 \right| < \varepsilon \right\},$$

satisfies $\lim_{n\to\infty} P^X(N_n) = 1$.

The proof is in Subsection 3.3.

Finally, we study a set of statistics. Let $\theta \in [-1,1]^M$. We suppose that $f(\cdot) = f(\cdot,\theta)$. The model (1) is then written in the form

$$Y_k = f(X_k,\theta) + \xi_k, \quad k = 1, \ldots, n,$$

and the vector $(X_1, Y_1, \ldots, X_n, Y_n)$ follows the law $\mathbb{P}_{f(\cdot,\theta)}$ that we will denote for brevity $\mathbb{P}_\theta$. For $X \in N_n$, consider the statistics

$$y_j = \frac{\sum_{k=1}^{n} Y_k f_j(X_k)}{\sum_{k=1}^{n} f_j^2(X_k)}, \quad j = 1, \ldots, M \tag{13}$$

where $f_j(x) = Lh^\beta\left(1 - \left|\frac{x-a_j}{h}\right|^\beta\right)_+$ for $x \in [0,1]$. For $X \in N_n$ the statistics y_j are well defined. These statistics satisfy the following proposition.

Proposition 5. *(i) For all $j \in \{1, \ldots, M\}$, the conditional distribution of y_j given $X \in N_n$ is gaussian with mean θ_j and variance v_j^2. The variance v_j^2 does not depend on θ and satisfies*

$$\frac{2\beta+1}{2\log(n)(1+\varepsilon)} \le v_j^2 \le \frac{2\beta+1}{2\log(n)(1-\varepsilon)}. \tag{14}$$

(ii) Conditionally on X, for $X \in N_n$, the variables y_j are independent.
(iii) In the model (1), with $f(\cdot) = f(\cdot,\theta)$, conditionally on X, for $X \in N_n$, $(y_1, \ldots, y_M)$ is a sufficient statistic for θ and the likelihood function of $(Y_1, \ldots, Y_n)$ conditionally on X, for $X \in N_n$, has the form

$$g(Y_1, \ldots, Y_n) = \prod_{i=1}^n \varphi_\sigma(Y_i) \prod_{j=1}^M \frac{\varphi_{v_j}(y_j - \theta_j)}{\varphi_{v_j}(y_j)},$$

where φ_v is the density of $\mathcal{N}(0, v^2)$ for $v > 0$.

The proof is in Subsection 3.3.

3.2.2 Proof of the inequality

Here we prove inequality (10). For $f \in \Sigma_Q(\beta, L)$ and an estimator θ_n, using the monotonicity of w and the Markov inequality we obtain that

$$\begin{aligned}\mathbb{E}_f\left[w\left(\psi_n^{-1}\|\theta_n - f\|_\infty\right)\right] \ge& w(C_0'(1-\varepsilon))\mathbb{P}_f\left[w\left(\psi_n^{-1}\|\theta_n - f\|_\infty\right) \ge w(C_0'(1-\varepsilon))\right] \\ \ge& w(C_0'(1-\varepsilon))\mathbb{P}_f\left[\psi_n^{-1}\|\theta_n - f\|_\infty \ge C_0'(1-\varepsilon)\right].\end{aligned}$$

Since $\Sigma' \subset \Sigma_Q(\beta, L)$ for n large enough, it is enough to prove that $\lim_{n\to\infty} \Lambda_n = 1$, where

$$\Lambda_n = \inf_{\theta_n} \sup_{f\in\Sigma'} \mathbb{P}_f\left(\psi_n^{-1}\|\theta_n - f\|_\infty \ge C_0'(1-\varepsilon)\right).$$

We have $\max_{j=1,\ldots,M} |\theta_n(a_j) - f(a_j)| \le \|\theta_n - f\|_\infty$. Setting $\hat\theta_j = \theta_n(a_j)C_0'\psi_n$ and using that $f(a_j) = C_0'\psi_n\theta_j$, we see that

$$\Lambda_n \ge \inf_{\hat\theta\in\mathbb{R}^M} \sup_{\theta\in[-1,1]^M} \mathbb{P}_\theta(C_n),$$

where $C_n = \{\max_{j=1,\ldots,M} |\hat\theta_j - \theta_j| \ge 1-\varepsilon\}$ and $\hat\theta = (\hat\theta_1, \ldots, \hat\theta_M) \in \mathbb{R}^M$ is measurable with respect to the (X_i, Y_i)'s. We have

$$\Lambda_n \ge \inf_{\hat\theta\in\mathbb{R}^M} \int_{\{-(1-\varepsilon),1-\varepsilon\}^M} \int_{N_n} \mathbb{P}_{\theta,X}(C_n)dP^X(X)\pi(d\theta), \tag{15}$$

where $\mathbb{P}_{\theta,X}$ is the distribution of $Y_1, \ldots, Y_n$ conditionally on $X = (X_1, \ldots, X_n)$ and π is the prior distribution on θ, $\pi(d\theta) = \prod_{j=1}^M \pi_j(d\theta_j)$, where π_j is the Bernoulli distribution on $\{-(1-\varepsilon), 1-\varepsilon\}$ that assigns probability 1/2 to $-(1-\varepsilon)$ and to $(1-\varepsilon)$. We will prove that for $X \in N_n$

$$H_n^X = \inf_{\hat\theta\in\mathbb{R}^M} \int \mathbb{P}_{\theta,X}(C_n)\pi(d\theta) \ge 1 + o(1), \tag{16}$$

where $o(1)$ is independent of X. This entails that

$$\int_{N_n} \inf_{\hat\theta\in\mathbb{R}^M} \int \mathbb{P}_{\theta,X}(C_n)\pi(d\theta)dP^X(X) \geq (1+o(1))P^X(N_n).$$

Using (15) and the Fubini and Fatou theorems, we find that Λ_n is greater than the left hand side of the last inequality. Thus we obtain that

$$\Lambda_n \geq P^X(N_n)(1+o(1)),$$

and by Lemma 2, we conclude that $\lim_{n\to\infty}\Lambda_n = 1$.

Proof of the inequality (16). We fix $X \in N_n$. We have

$$H_n^X = 1 - \sup_{\hat\theta\in\mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbb{P}_{\theta,X}\pi(d\theta), \tag{17}$$

where the supremum is taken over all the estimators $\hat\theta$ measurable with respect to the $(X_i, Y_i)'s$. By Proposition 5, the Fubini and Fatou theorems we have

$$\sup_{\hat\theta\in\mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbb{P}_{\theta,X}\pi(d\theta)$$

$$\leq \int \frac{\prod_{i=1}^n \varphi_\sigma(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left(\sup_{\hat\theta\in\mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} \varphi_{v_j}(y_j-\theta_j)\pi_j(d\theta_j) \right) dY_1\cdots dY_n$$

$$= \int \frac{\prod_{i=1}^n \varphi_\sigma(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left(\prod_{j=1}^M \sup_{\hat\theta_j\in\mathbb{R}} \int I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} \varphi_{v_j}(y_j-\theta_j)\pi_j(d\theta_j) \right) dY_1\cdots dY_n.$$

It is not hard to prove that the problem of maximization

$$\max_{\hat\theta_j\in\mathbb{R}} \int I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} \varphi_{v_j}(y_j-\theta_j)\pi_j(d\theta_j)$$

has the solution $\hat\theta_j(y_j) = (1-\varepsilon)I_{\{y_j\geq 0\}} - (1-\varepsilon)I_{\{y_j<0\}}$. This allows to compute the supremum in (17) directly but to avoid calculations we can argue in follows. The above inequality is in fact an equality and we have

$$\sup_{\hat\theta\in\mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbb{P}_{\theta,X}\pi(d\theta) = \max_{\hat\theta\in\Upsilon} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbb{P}_{\theta,X}\pi(d\theta),$$

where the max is taken over the class Υ of all the estimators of the form $\hat\theta = (\hat\theta_1(y_1), \ldots, \hat\theta_M(y_M))$ where $\hat\theta_j$ is a measurable function of y_j with values in $\{-(1-\varepsilon), 1-\varepsilon\}$ and the supremum is taken on the estimators which are measurable with respect to the (X_i, Y_i)'s. Moreover we have, as $\hat\theta$ depends only on $T = (y_1, \ldots, y_M)$

$$\max_{\hat\theta\in\Upsilon} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbb{P}_{\theta,X}\pi(d\theta) = \max_{\hat\theta\in\Upsilon} \int \prod_{j=1}^M I_{\{|\hat\theta_j-\theta_j|<1-\varepsilon\}} d\mathbf{P}_{T,X}\pi(d\theta), \tag{18}$$

where $\mathbf{P}_{T,X}$ is the probability associated to the statistic $T = (y_1, \dots, y_M)$ conditionally on X. The quantity (18) is also equal by Proposition 5 to

$$\prod_{j=1}^{M} \max_{\hat\theta_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat\theta_j(y_j)-\theta_j|<1-\varepsilon\}} \varphi_{v_j}(y_j-\theta_j) dy_j \pi_j(d\theta_j),$$

and then we have

$$H_n^X \geq 1 - \prod_{j=1}^{M} \max_{\hat\theta_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat\theta_j(y_j)-\theta_j|<1-\varepsilon\}} \varphi_{v_j}(y_j-\theta_j) dy_j \pi_j(d\theta_j) \tag{19}$$

$$= 1 - \prod_{j=1}^{M} \left(1 - r_j^X\right),$$

where

$$r_j^X = \min_{\hat\theta_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat\theta_j(u)-\theta_j|\geq 1-\varepsilon\}} \varphi_{v_j}(u-\theta_j) du \pi_j(d\theta_j).$$

We denote by $P^j_{\theta,X}$ the probability measure with density $u \to \varphi_{v_j}(u-\theta)$. We have

$$\begin{aligned}
r_j^X =& \frac{1}{2} \min_{\hat\theta_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \left[P^j_{(1-\varepsilon),X} \left\{ |\hat\theta_j - (1-\varepsilon)| \geq (1-\varepsilon) \right\} + P^j_{-(1-\varepsilon),X} \left\{ |\hat\theta_j + (1-\varepsilon)| \geq (1-\varepsilon) \right\} \right] \\
\geq& \frac{1}{2} \min_{\hat\theta_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \left[\int (I_{\{\hat\theta_j \leq 0\}} + I_{\{\hat\theta_j \geq 0\}}) \min\left(dP^j_{(1-\varepsilon),X}, dP^j_{-(1-\varepsilon),X} \right) \right] \\
=& \frac{1}{2} \int \min\left(dP^j_{(1-\varepsilon),X}, dP^j_{-(1-\varepsilon),X} \right) \\
=& \frac{1}{v_j} \int_{-\infty}^{0} \varphi_1\left(\frac{y-(1-\varepsilon)}{v_j} \right) dy = \Phi\left(-\frac{(1-\varepsilon)}{v_j} \right),
\end{aligned}$$

where Φ is the standard normal cdf. By the inequality (14) and using that $\Phi(-z) = \frac{1}{z\sqrt{2\pi}} \exp(-z^2/2)(1+o(1))$ for $z \to +\infty$, we have

$$\Phi\left(-\frac{(1-\varepsilon)}{v_j} \right) = \frac{v_j\sqrt{2\pi}}{(1-\varepsilon)\sqrt{2\pi}} \exp\left(-\frac{(1-\varepsilon)^2}{2v_j^2} \right)(1+o(1))$$

as $n \to \infty$ and using inequality (14) for v_j, we get

$$\inf_{X \in N_n} r_j^X \geq \frac{D_4}{\sqrt{\log n}} n^{-\frac{(1-\varepsilon)^2(1+\varepsilon)}{2\beta+1}} (1+o(1)).$$

Now $M = O\left(\left(\frac{n}{\log n} \right)^{\frac{1-\varepsilon}{2\beta+1}} \right)$, therefore

$$\inf_{X \in N_n} M r_j^X \geq D_5 (\log n)^{\frac{1-\varepsilon}{2\beta+1} - \frac{1}{2}} n^{\frac{(1-\varepsilon)\varepsilon^2}{2\beta+1}} (1+o(1)).$$

From this last inequality and inequality (19), we obtain the inequality (16), which finishes the proof of the lower bound.

3.3 Proofs of lemmas and propositions

3.3.1 Proof of Lemma 1

We are going to prove that the event A_n satisfies

$$P^X(A_n) \geq 1 - 2\frac{n}{m}\exp\left(-c_A n h \delta_n^2\right),$$

for a constant $c_A > 0$. There are similar results for the events $A_{1,n}, \ldots, A'_{2,n}$ with other constants. Together these results entail the lemma. We are going to use Bernstein's inequality. First we take a point $x_k \in [h, 1-h]$. The proof will be similar for $x_k \in [0,h) \cup (1-h, 1]$ and we define the random variables Z_i, for $i \in \{1, \ldots, n\}$ by

$$Z_i = \frac{1}{h}K\left(\frac{X_i - x_k}{h}\right) - \mathbb{E}_f\left[\frac{1}{h}K\left(\frac{X_i - x_k}{h}\right)\right].$$

These variables satisfy $\mathbb{E}_f[Z_i] = 0$, $\mathbb{E}_f\left[Z_i^2\right] \leq \frac{K_{max}^2 \mu_1}{h}$ and $|Z_i| \leq \frac{2K_{max}}{h}$. The constant K_{max} is such that $K(x) \leq K_{max}$ for all x in $[-1,1]$ and μ_1 is such that $\mu(x) \leq \mu_1$ for all x in $[0,1]$ (such μ_1 exists because μ is continuous). Let $A(k) = \left\{\left|\mu(x_k) - \frac{1}{nh}\sum_{i=1}^n K\left(\frac{X_i - x_k}{h}\right)\right| < \delta_n\right\}$. We have

$$P^X(A(k)) = P^X\left(\left|\frac{1}{n}\sum_{i=1}^n (Z_i + \delta_{k,h})\right| < \delta_n\right),$$

where $\delta_{k,h} = \int_{-1}^1 K(y)[\mu(x_k + yh) - \mu(x_k)]dy$. As μ satisfies a Lipschitz condition, $\delta_{k,h}$ satisfies $|\delta_{k,h}| \leq \rho h \int_{-1}^1 K(y)|y|dy$ with $\rho > 0$. We have for n large enough

$$P^X(A(k)) \geq P^X\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| < \delta_n - |\delta_{k,h}|\right).$$

By Bernstein's inequality applied to the variables Z_i, we have

$$P^X\left(\left|\frac{1}{n}\sum_{i=1}^n Z_i\right| < \delta_n - |\delta_{k,h}|\right) \geq 1 - 2\exp\left(-\frac{n(\delta_n - |\delta_{k,h}|)^2}{2\left(\frac{K_{max}^2\mu_1}{h} + \frac{2(\delta_n - |\delta_{k,h}|)K_{max}}{3h}\right)}\right).$$

Using the fact that $\delta_{k,h} = O(h)$, we obtain that for n large enough, there exists a constant c_A independent of k such that

$$P^X(A(k)) \geq 1 - 2\exp\left(-c_A n h \delta_n^2\right).$$

From this we deduce easily the result about A_n because $card\{k\} \leq \frac{n}{m}$.

3.3.2 Proof of Proposition 1

Let $f \in \Sigma_Q(\beta, L)$. We have

$$\begin{aligned}
\mathbb{E}_f\left[w\left(\psi_n^{-1}\|f - \theta_n^*\|_\infty\right) I_{B_n^C}\right] &\leq \sqrt{\mathbb{E}_f\left[w^2\left(\psi_n^{-1}\|f - \theta_n^*\|_\infty\right)\right]}\sqrt{\mathbb{P}_f(B_n^C)} \\
&\leq \sqrt{\mathbb{E}_f\left(1 + \left(\psi_n^{-1}\|f - \theta_n^*\|_\infty\right)^\gamma\right)^2}\sqrt{P^X(B_n^C)}
\end{aligned}$$

since the event B_n only depends on X,

$$\leq \sqrt{2}\sqrt{1+\mathbb{E}_f\left(\left(\psi_n^{-1}\|f-\theta_n^*\|_\infty\right)^{2\gamma}\right)}\sqrt{P^X(B_n^C)}.$$

Now $\mathbb{E}_f\left(\left(\psi_n^{-1}\|f-\theta_n^*\|_\infty\right)^{2\gamma}\right) \leq \psi_n^{-2\gamma}D_6\left(Q^{2\gamma}+\mathbb{E}_f\|\theta_n^*\|_\infty^{2\gamma}\right)$. Some algebra and the fact that

$$\max\left(\frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{X_j-x_k}{h}\right),\delta_n\right) \geq \delta_n,$$

yield $\mathbb{E}_f\|\theta_n^*\|_\infty^{2\gamma} = O(n^{\gamma_1})$, with some $\gamma_1 \geq 0$. From the relations above and Lemma 1, we deduce that $\lim_{n\to\infty}\mathbb{E}_f\left[w\left(\psi_n^{-1}\|f-\theta_n^*\|_\infty\right)I_{B_n^C}\right] = 0$.

3.3.3 Proof of Proposition 2

Let $f \in \Sigma_Q(\beta,L)$ and $x_k \in [h,1-h]$. Consider n large enough such that $\delta_n \leq \mu(x_k)-\delta_n$. We have on B_n

$$\mu(x_k)-\delta_n \leq \frac{1}{nh}\sum_{j=1}^{n}K\left(\frac{X_j-x_k}{h}\right).$$

Thus some algebra and the fact $f \in \Sigma_Q(\beta,L)$ yield

$$\begin{aligned}\left|\mathbb{E}_f(\theta_n^*(x_k)I_{B_n})-f(x_k)P^X(B_n)\right| &\leq \left|\frac{1}{\mu(x_k)-\delta_n}\mathbb{E}_f\left[\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{X_i-x_k}{h}\right)(f(X_i)-f(x_k))I_{B_n}\right]\right| \\ &\leq \frac{Lh^\beta\int_{-1}^{1}|y|^\beta K(y)\mu(x_k+yh)dy}{\mu(x_k)-\delta_n} \\ &\leq \frac{Lh^\beta\mu(x_k)(1+o(1))}{(\mu(x_k)-\delta_n)(2\beta+1)}.\end{aligned}$$

For x_k belonging to $[0,h)$ or $(1-h,1]$, we have the same result. Thus for all $k \in \{1,\ldots,[\frac{n}{m}]\}$

$$\psi_n^{-1}|\mathbb{E}_f(\theta_n^*(x_k)I_{B_n})-f(x_k)P^X(B_n)| \leq \left(1+\frac{\delta_n}{\mu_0-\delta_n}\right)\frac{C_0'(1+o(1))}{2\beta+1} \leq \frac{C_0'(1+o(1))}{2\beta+1}.$$

As $f \in \Sigma_Q(\beta,L)$, we have for $x \in [0,1]$

$$\begin{aligned}|b_n(x,f)| &\leq \max_{k\in\{1,\ldots,[\frac{n}{m}]\}}|b_n(x_k,f)| + L\left(\frac{m}{n}\right)^\beta P^X(B_n) \\ &\leq \max_{k\in\{1,\ldots,[\frac{n}{m}]\}}|b_n(x_k,f)| + \psi_n\delta_n^\beta L(1+o(1)).\end{aligned}$$

Then we obtain that $\psi_n^{-1}|b_n(x,f)| \leq \frac{C_0'(1+o(1))}{2\beta+1}$ with $o(1)$ independent of f.

3.3.4 Proof of Proposition 3

Let $f \in \Sigma_Q(\beta, L)$, $z > 1$ and

$$P_n = \mathbb{P}_f\left[\left\{\max_k \psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right].$$

We have

$$P_n \leq \sum_k \mathbb{P}_f\left[\left\{\psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right].$$

We are going to reason for $x_k \in [h, 1-h]$, but the proof is similar for $x_k \in [0, h) \cup (1-h, 1]$. As B_n depends only on $X_1, \ldots, X_n$, we have

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right] = \mathbb{E}_f\left[\mathbb{P}_f\left[\psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1} \,\Big|\, X_1, \ldots, X_n\right] I_{B_n}\right].$$

The variable $\widehat{Z}_n(x_k, f)$ is gaussian conditionally on the X_i's, with conditional variance equal to

$$\frac{\sigma^2 (P^X(B_n))^2 \sum_{j=1}^n K^2\left(\frac{X_j - x_k}{h}\right)}{\left(\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)\right)^2}.$$

Since n has been chosen such that $P^X(B_n) > 0$ in the definition of the stochastic term, we obtain

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right] \leq \mathbb{E}_f\left[\exp\left[-\frac{\psi_n^2 (2\beta C_0' z)^2 \left(\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)\right)^2}{\sigma^2 (P^X(B_n))^2 \sum_{j=1}^n K^2\left(\frac{X_j - x_k}{h}\right)}\right] I_{B_n}\right].$$

Replacing the expression for h and σ^2 in terms of n, C_0', L, β and μ_0, we obtain that the quantity above in the right hand side is equal to

$$\mathbb{E}_f\left[\exp\left[-\frac{z^2 \log n (\beta+1)\left(\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)\right)^2}{\mu_0 (P^X(B_n))^2 (2\beta+1)^2 \frac{1}{nh}\sum_{j=1}^n K^2\left(\frac{X_j - x_k}{h}\right)}\right] I_{B_n}\right].$$

We have on B_n

$$\sum_{j=1}^n \frac{1}{nh} K\left(\frac{X_j - x_k}{h}\right) \geq \mu(x_k) - \delta_n \geq \mu_0 - \delta_n,$$

$$\sum_{j=1}^n \frac{1}{nh} K^2\left(\frac{X_j - x_k}{h}\right) \leq \frac{\beta+1}{2\beta+1}\mu(x_k) + \delta_n.$$

Consider n large enough such that $\mu_0 - \delta_n > 0$. Thus we deduce using the inequalities above that

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}|\widehat{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right] \leq \exp\left[-\frac{z^2 \log n C(\delta_n)}{(2\beta+1)}\right],$$

with

$$C(\delta_n) = \frac{(\mu_0 - \delta_n)}{\mu_0 (P^X(B_n))^2}\left[1 - \frac{\delta_n (3\beta+2)}{(\beta+1)\left(\mu_1 + \frac{\delta_n (2\beta+1)}{\beta+1}\right)}\right].$$

The quantity $C(\delta_n)$ tends to 1 as $n \to \infty$. Because of the fact that $card\{k\} \leq \delta_n^{-1}\psi_n^{-1/\beta}$, we have

$$P_n \leq \delta_n^{-1}(\log n)^{-\frac{1}{2\beta+1}} n^{-\alpha_1(n)}.$$

3.3.5 Proof of Proposition 4

Let $f \in \Sigma_Q(\beta, L)$. We are still going to reason for $x_k \in [h, 1-h]$, and the proof is similar for $x_k \in [0, h) \cup (1-h, 1]$. Let $\widetilde{U}_n(x_k, f) = U_n(x_k, f)P^X(B_n)I_{B_n} - \mathbb{E}_f[U_n(x_k, f)P^X(B_n)I_{B_n}]$. If B_n holds, we have $\widetilde{Z}_n(x_k, f) = \widetilde{U}_n(x_k, f) - \mathbb{E}_f[U_n(x_k, f)I_{B_n}]P^X(B_n^C)$. Consider n large enough such that for all $z \geq \varepsilon/2$ we have

$$|\mathbb{E}_f[U_n(x_k, f)I_{B_n}]P^X(B_n^C)| \leq \frac{\beta C_0' z\psi_n}{2\beta+1}.$$

Such choice of n is possible in view of Lemma 1, since $\mathbb{E}_f[U_n(x_k, f)I_{B_n}]$ is bounded. Thus we have

$$\mathbb{P}_f\left[\left\{\psi_n^{-1}|\widetilde{Z}_n(x_k, f)| > \frac{2\beta C_0' z}{2\beta+1}\right\} \cap B_n\right] \leq \mathbb{P}_f\left[\psi_n^{-1}|\widetilde{U}_n(x_k, f)| > \frac{\beta C_0' z}{2\beta+1}\right].$$

We are going to apply Bernstein's inequality to the variable $\widetilde{U}_n(x_k, f)$ which is a zero-mean variable bounded by $2Q$. Since $\mu(x_k) \geq \mu_0$, the variance of $\widetilde{U}_n(x_k, f)$ satisfies

$$\begin{aligned}
\mathbb{E}_f\left[\widetilde{U}_n(x_k, f)^2\right] \leq & \mathbb{E}_f\left[\left(\frac{\frac{1}{nh}\sum_{j=1}^n f(X_j)K\left(\frac{X_j - x_k}{h}\right)}{\frac{1}{nh}\sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)}\right)^2 I_{B_n}(P^X(B_n))^2\right], \\
\leq & \frac{1}{(\mu(x_k) - \delta_n)^2 nh^2}\mathbb{E}_f\left[f^2(X_1)K^2\left(\frac{X_1 - x_k}{h}\right)\right](P^X(B_n))^2, \\
\leq & \frac{Q^2\mu(x_k)(1+o(1))}{(\mu(x_k) - \delta_n)^2 nh}\int_{-1}^1 K^2(y)dy, \\
\leq & \frac{D_7(1+o(1))}{nh},
\end{aligned}$$

where $o(1)$ is uniform in $f \in \Sigma_Q(\beta, L)$ and D_7 is independent of $f \in \Sigma_Q(\beta, L)$, n and k. By applying Bernstein's inequality to the variable $\widetilde{U}_n(x_k, f)$ (note that here the family of random variables contains only one summand), we obtain

$$\mathbb{P}_f\left[\psi_n^{-1}|\widetilde{U}_n(x_k, f)| > \frac{\beta C_0' z}{2\beta+1}\right] \leq 2\exp\left(-\frac{\lambda^2}{2\left(\frac{D_7(1+o(1))}{nh} + \frac{2\lambda Q}{3}\right)}\right),$$

where $\lambda = \frac{\psi_n \beta C_0' z}{2\beta+1}$. Thus for n large enough, we have

$$\mathbb{P}_f\left[\psi_n^{-1}|\widetilde{U}_n(x_k, f)| > \frac{\beta C_0' z}{2\beta+1}\right] \leq 2\exp\left(-D_0 z\psi_n\right),$$

with D_0 independent of $f \in \Sigma_Q(\beta, L)$ and k. To finish the proof, it is enough to note that $card\{k\} = [\frac{n}{m}] \leq \delta_n^{-1}\psi_n^{-1/\beta}$.

3.3.6 Proof of Lemma 2

Like in Lemma 1, using Bernstein's inequality we obtain that for n large enough

$$P^X\left\{\Big|\frac{(\beta+1)(2\beta+1)}{4\mu_0\beta^2 nh}\sum_{k=1}^{n}(1-|\frac{X_k-a_j}{h}|^{\beta})_+^2-1\Big|\geq\varepsilon\right\}\leq 2\exp(-nhD_8)$$

where D_8 is a constant which depends on ε, but does not depend on n. Now

$$P^X\left[N_n^C\right]\leq\sum_{j=0}^{M}P^X\left\{\Big|\frac{(\beta+1)(2\beta+1)}{4\mu_0\beta^2 nh}\sum_{k=1}^{n}(1-|\frac{X_k-a_j}{h}|^{\beta})_+^2-1\Big|\geq\varepsilon\right\}.$$

Thus

$$P^X\left[N_n^C\right]\leq 2M\exp(-nhD_8).$$

Since $M=O\left(\left(\frac{n}{\log n}\right)^{\frac{1-\varepsilon}{2\beta+1}}\right)$, we deduce that $\lim_{n\to\infty}P^X\left[N_n^C\right]=0$.

3.3.7 Proof of Proposition 5

(i) The fact that y_j is conditionally gaussian with conditional mean θ_j comes from the definition of y_j and the fact that the functions f_j have disjoint supports. The conditional variance of y_j satisfies for $X\in N_n$

$$\begin{aligned}Var(y_j|X)=&\mathbb{E}_f\left[\frac{\left(\sum_{k=1}^n\xi_k f_j(X_k)\right)^2}{\left(\sum_{k=1}^n f_j^2(X_k)\right)^2}|X\right]\\ =&\frac{\sigma^2}{\sum_{k=1}^n f_j^2(X_k)}\\ =&\frac{\sigma^2}{L^2h^{2\beta}\sum_{k=1}^n\left(1-\left|\frac{X_k-a_j}{h}\right|^{\beta}\right)_+^2}.\end{aligned}$$

Using that $X\in N_n$ and replacing the expression for σ and h in terms of C_0', n, β and L, we obtain the inequality for v_j.
(ii) comes from the fact the functions f_j have disjoint supports and that the ξ_i's are independent and independent of the X_i's. (iii) is obtained by calculating the likelihood function of $Y_1,\ldots,Y_n$ conditionally on X, for $X\in N_n$.

4 Appendix

Proposition 6. *For all $\theta\in[-1,1]^M$, we have $f(\cdot,\theta)\in\Sigma(\beta,L)$.*

Proof. Let $x \in [0,1]$. We set $f_j(x) = Lh^\beta \left(1 - \left|\frac{x-a_j}{h}\right|^\beta\right)_+$.

Then we have $f(x,\theta) = \sum_{i=0}^{M} \theta_j f_j(x)$.

- First we are going to show that $f_j \in \Sigma(\beta, L)$. Let $x, x' \in [0,1]$. We have

$$\begin{aligned}
\left|f_j(x) - f_j(x')\right| &\leq Lh^\beta \left| \left|\frac{x-a_j}{h}\right|^\beta - \left|\frac{x'-a_j}{h}\right|^\beta \right| \\
&\leq Lh^\beta \left|\frac{x-a_j}{h} - \frac{x'-a_j}{h}\right|^\beta \\
&\leq L\left|x - x'\right|^\beta.
\end{aligned}$$

This gives the result.

- now, let show that $f(\cdot,\theta) \in \Sigma(\beta, L)$. We consider the case $x \in [a_j - h, a_j + h]$ and $x' \in [a_{j+1} - h, a_{j+1} + h]$, for $j \in \{1, \ldots, M-1\}$. We have

$$\begin{aligned}
|f(x) - f(x')| &= \left|\theta_j f_j(x) - \theta_{j+1} f_{j+1}(x')\right| \\
&\leq |f_j(x) - f_j(a_j + h)| + \left|f_{j+1}(a_{j+1} - h) - f_{j+1}(x')\right| \\
&\leq 2L(2h)^\beta \\
&\leq L\left(2^{1/\beta} 2h\right)^\beta.
\end{aligned}$$

Since $|x - x'| \geq 2^{1/\beta} 2h$ we obtain that $|f(x) - f(x')| \leq |x - x'|^\beta$. From this particular case we can easily show that $|f(x) - f(x')| \leq |x - x'|^\beta$ for all $x, x' \in [0,1]$. □

References

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: solution via optimal recovery. *Probab. Theory Related Fields*, 99(2):145–170.

Efromovich, S. (1999). *Nonparametric curve estimation*. Springer Series in Statistics. Springer-Verlag, New York. Methods, theory, and applications.

Ibragimov, I. A. and Hasminskii, R. Z. (1981). *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York. Asymptotic theory, Translated from the Russian by Samuel Kotz.

Ibragimov, I. A. and Hasminskii, R. Z. (1982). Bounds for the quality of nonparametric estimation of regression. *Theory Probab. Appl.*, 27(1):81–94.

Korostelev, A. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. *Theory Probab. Appl.*, 38(4):875–882.

Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, 5(6):1099–1118.

Leonov, S. L. (1997). On the solution of an optimal recovery problem and its applications in nonparametric regression. *Math. Methods Statist.*, 6(4):476–490 (1998).

Lepski, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, volume 12 of *Adv. Soviet Math.*, pages 87–106. Amer. Math. Soc., Providence, RI.

Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1):17–48.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4):1040–1053.

Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin.