

# 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  $f$  defined on  $[0, 1]$  from observations

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

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

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

where  $m = \lfloor \beta \rfloor$  is an integer such that  $0 < \alpha \leq 1$  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

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

and  $\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|$ . We suppose that the  $X_i$  have a density  $\mu$  w.r.t. the Lebesgue measure,  $\mu$  belongs to a Hölder class  $\Sigma(l, C)$  with  $0 < l \leq 1$  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  $\theta_n$  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: kbertain@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  $W_0, \gamma$  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,  $\psi_n$  is the minimax rate of convergence (c.f. Ibragimov and Has'minskii (1981,1982), Stone (1982)).

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

$$C = \lim_{n \rightarrow \infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\theta_n^*), \quad (3)$$

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

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

Here  $h$  is a bandwidth depending on  $n$  and

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

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  $g_\beta$  of the optimization problem

$$\max g_\beta(0) \quad \text{subject to} \quad \begin{cases} \|g_\beta\|_2 \leq 1 \\ g_\beta \in \Sigma(\beta, 1), \end{cases} \quad (6)$$

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  $g_\beta$ :

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

and that the exact constant depends on  $g_\beta(0)$ . 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  $g_\beta$  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 Efremovich (1999) and Tsybakov (2003).

Our results are the following. In Section 2, we give an asymptotically exact estimator  $\theta_n^*$  and the exact constant for the regression model with random design. If the density  $\mu$  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  $\theta_n^*$  depend on the minimum value of the design density  $\mu_0$ . It means that the asymptotically minimax estimators contribute to the sup-norm risk essentially at the points where we have less observations. The estimator  $\theta_n^*$  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  $\theta_n^*$ . We shall prove in Subsection 3.1 that  $\theta_n^*$  is an asymptotically exact estimator. This estimator is close to a Nadaraya-Watson estimator with the kernel  $K$  defined in (5). The bandwidth of  $\theta_n^*$  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  $\theta_n^*$  in a regular grid of points  $x_k = \frac{km}{n} \in [0, 1]$  for  $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$ , with  $m = \lfloor \delta_n n \psi_n^{\frac{1}{\beta}} + 1 \rfloor$ ,  $\delta_n = \frac{1}{\log n}$  and  $\lfloor x \rfloor$  denotes the integer part of  $x$ . To account for the boundary effects, we need to introduce other kernels:

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

The estimator  $\theta_n^*$  is defined for  $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$  by

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

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  $\theta_n^*$  is defined to be the polygonal function connecting the points  $(x_k, \theta_n^*(x_k))$  for  $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$ . Moreover, we put  $\theta_n^*(x) = \theta_n^*(x_1)$  if  $x \in [0, x_1]$  and if  $x_{\lfloor \frac{n}{m} \rfloor} < 1$  we put  $\theta_n^*(x) = \theta_n^*(x_{\lfloor \frac{n}{m} \rfloor})$  for  $x \in [x_{\lfloor \frac{n}{m} \rfloor}, 1]$ .

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  $\theta_n^*$  satisfies*

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

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  $\theta_n^*$  satisfies

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

In Subsection 3.2, we prove that

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

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  $\delta_n$  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  $\infty$ , 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  $\theta_n^*$  does not depend on  $Q$ , but it depends on  $\mu_0$ . It is possible to construct an asymptotically exact estimator independent of  $\mu_0$  and  $Q$  but the proof is rather technical. For this purpose, we cut the sample  $(X_1, \dots, X_n)$  in two parts of size  $\alpha_n$  and  $n - \alpha_n$ , where  $\alpha_n$  is an integer such that  $\alpha_n \rightarrow \infty$  and  $\alpha_n/n \rightarrow 0$  as  $n \rightarrow \infty$ . We estimate  $\mu_0$  with the part  $(X_1, \dots, X_{\alpha_n})$  of the sample by

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

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  $g_n \rightarrow 0$  and  $\alpha_n g_n \rightarrow \infty$ . We construct an estimator of  $f$  in the same way as  $\theta_n^*$  except we only use the part  $(X_{\alpha_n+1}, \dots, X_n)$  of the sample and we replace  $\mu_0$  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  $(X_1, \dots, X_{\alpha_n})$ .

(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  $g_\beta$  of the problem (6) is not explicitly known except for  $\beta = 2$ ). A similar result could be found and the exact constant will be  $w(C_1)$  with

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

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

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

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  $g_\beta$  and the fact that  $g_\beta$  is compactly supported. This was proved by Leonov (1997). He also proved that  $g_\beta$  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  $\sigma^{-2}$  serves to replace the density of the design points. In this model, the maximum of  $\sigma^2$  corresponds to the minimum value of the design density  $\mu_0$ . 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  $\beta$  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*

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

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,  $D_0, D_1, \dots$  are positive constants.

**Proposition 1.** *We have*

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

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

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

**Proposition 2.** *The bias term satisfies*

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

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

$$\begin{aligned} \hat{Z}_n(x_k, f) &= \frac{\frac{1}{nh} \sum_{j=1}^n \xi_j K\left(\frac{X_j - x_k}{h}\right)}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)} P^X(B_n), \\ \tilde{Z}_n(x_k, f) &= U_n(x_k, f) P^X(B_n) - \mathbb{E}_f(U_n(x_k, f) I_{B_n}), \end{aligned}$$

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  $(1-h, 1]$ ),  $\hat{Z}_n(x_k, f)$ ,  $\tilde{Z}_n(x_k, f)$  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  $\hat{Z}_n(\cdot, f)$  satisfies for all  $z > 1$  and  $n$  large enough*

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

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

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

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

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

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

$$\begin{aligned} \mathbb{E}_f(w(\Delta_n)I_{B_n}) &\leq w(C_0(1+\varepsilon)^2)\mathbb{P}_f[\Delta_n I_{B_n} \leq C_0(1+\varepsilon)^2] \\ &\quad + (\mathbb{E}_f(w^2(\Delta_n)I_{B_n}))^{\frac{1}{2}} (\mathbb{P}_f[\Delta_n I_{B_n} > C_0(1+\varepsilon)^2])^{\frac{1}{2}}. \end{aligned}$$

To obtain the inequality (9), it is enough to prove that

$$(i) \lim_{n \rightarrow \infty} \sup_{f \in \Sigma_Q(\beta, L)} \mathbb{P}_f[\Delta_n I_{B_n} > C_0(1+\varepsilon)^2] = 0,$$

$$(ii) \text{ there exists a constant } D_1 \text{ such that } \limsup_{n \rightarrow \infty} \sup_{f \in \Sigma_Q(\beta, L)} \mathbb{E}_f(w^2(\Delta_n)I_{B_n}) \leq D_1.$$

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

$$\begin{aligned} \mathbb{P}_f[\Delta_n I_{B_n} > C_0(1+\varepsilon)^2] &= \mathbb{P}_f[\Delta_n I_{B_n} P^X(B_n) > C_0 P^X(B_n)(1+\varepsilon)^2] \\ &\leq \mathbb{P}_f[\Delta_n I_{B_n} P^X(B_n) > C_0(1+\varepsilon)]. \end{aligned}$$

Note also that

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

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

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

Since  $\theta_n^*$  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

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

Since  $C(\delta_n)$  tends to 1 as  $n \rightarrow \infty$ , 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  $\infty$  uniformly in  $f \in \Sigma_Q(\beta, L)$ . So we obtain (i).

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

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

Using the fact that

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

Propositions 3 and 4, and noting that  $C(\delta_n)$  tends to 1 as  $n \rightarrow \infty$ , we prove that

$$\limsup_{n \rightarrow \infty} \mathbb{E}_f\left[(\psi_n^{-1}\|Z_n(\cdot, f)\|_\infty)^{2\gamma}\right] < \infty. \text{ This and Proposition 2 entail (ii).}$$

## 3.2 Proof of inequality (10)

### 3.2.1 Preliminaries

First, we need to define  $\Sigma'$ , a subspace of  $\Sigma_Q(\beta, L)$ . Since  $\mu$  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  $\varepsilon$  introduced in Section 2. Let  $M = \left\lceil \frac{\gamma_n}{2h(2^{1/\beta}+1)} \right\rceil$  and define the points  $a_1, \dots, a_M$  in a neighbourhood of  $x_0$  in the following way. For  $n$  large enough and if  $x_0 \in (0, 1)$ , we put

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

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

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

where for  $\theta = (\theta_1, \dots, \theta_M) \in [-1, 1]^M$  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  $\theta \in [-1, 1]^M$ ,  $f(\cdot, \theta) \in \Sigma(\beta, L)$  and  $\|f\|_\infty \leq Q$  for  $n$  large enough. Therefore for  $n$  large enough  $\Sigma' \subset \Sigma_Q(\beta, L)$ .

**Remark:** For  $\beta > 1$  the subspace  $\Sigma'$  should be defined in a similar way:

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

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

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

**Lemma 2.** *The event*

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

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

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

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

and the vector  $(X_1, Y_1, \dots, X_n, Y_n)$  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

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



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  $j \in \{1, \dots, M\}$ , the conditional distribution of  $y_j$  given  $X \in N_n$  is gaussian with mean  $\theta_j$  and variance  $v_j^2$ . The variance  $v_j^2$  does not depend on  $\theta$  and satisfies

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

(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  $X \in N_n$ ,  $(y_1, \dots, y_M)$  is a sufficient statistic for  $\theta$  and the likelihood function of  $(Y_1, \dots, Y_n)$  conditionally on  $X$ , for  $X \in N_n$ , has the form

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

where  $\varphi_v$  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  $\theta_n$ , using the monotonicity of  $w$  and the Markov inequality we obtain that

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

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

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

We have  $\max_{j=1, \dots, M} |\theta_n(a_j) - f(a_j)| \leq \|\theta_n - f\|_\infty$ . 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

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

where  $C_n = \{\max_{j=1, \dots, M} |\hat{\theta}_j - \theta_j| \geq 1 - \varepsilon\}$  and  $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_M) \in \mathbb{R}^M$  is measurable with respect to the  $(X_i, Y_i)$ 's. We have

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

where  $\mathbb{P}_{\theta, X}$  is the distribution of  $Y_1, \dots, Y_n$  conditionally on  $X = (X_1, \dots, X_n)$  and  $\pi$  is the prior distribution on  $\theta$ ,  $\pi(d\theta) = \prod_{j=1}^M \pi_j(d\theta_j)$ , where  $\pi_j$  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  $X \in N_n$

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

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  $\Lambda_n$  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  $\lim_{n \rightarrow \infty} \Lambda_n = 1$ .

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

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

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

$$\begin{aligned} & \sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1 - \varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta) \\ & \leq \int \frac{\prod_{i=1}^n \varphi_{\sigma}(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left( \sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1 - \varepsilon\}} \varphi_{v_j}(y_j - \theta_j) \pi_j(d\theta_j) \right) dY_1 \cdots dY_n \\ & = \int \frac{\prod_{i=1}^n \varphi_{\sigma}(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left( \prod_{j=1}^M \sup_{\hat{\theta}_j \in \mathbb{R}} \int I_{\{|\hat{\theta}_j - \theta_j| < 1 - \varepsilon\}} \varphi_{v_j}(y_j - \theta_j) \pi_j(d\theta_j) \right) dY_1 \cdots dY_n. \end{aligned}$$

It is not hard to prove that the problem of maximisation

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

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  $\Upsilon$  of all the estimators of the form  $\hat{\theta} = (\hat{\theta}_1(y_1), \dots, \hat{\theta}_M(y_M))$  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  $T = (y_1, \dots, y_M)$

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

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

$$\begin{aligned} H_n^X &\geq 1 - \prod_{j=1}^M \max_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat{\theta}_j(y_j) - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) dy_j \pi_j(d\theta_j) \\ &= 1 - \prod_{j=1}^M (1 - r_j^X), \end{aligned} \quad (19)$$

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  $P_{\hat{\theta}_j, X}^j$  the probability measure with density  $u \rightarrow \varphi_{v_j}(u - \theta)$ . We have

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

where  $\Phi$  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  $z \rightarrow +\infty$ , 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  $n \rightarrow \infty$  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

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

for a constant  $c_A > 0$ . There are similar results for the events  $A_{1,n}, \dots, A'_{2,n}$  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  $i \in \{1, \dots, n\}$  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  $\mathbb{E}_f[Z_i] = 0$ ,  $\mathbb{E}_f[Z_i^2] \leq \frac{K_{max}^2 \mu_1}{h}$  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  $\mu_1$  is such that  $\mu(x) \leq \mu_1$  for all  $x$  in  $[0, 1]$  (such  $\mu_1$  exists because  $\mu$  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  $\mu$  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

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

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

#### 3.3.2 Proof of Proposition 1

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

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

since the event  $B_n$  only depends on  $X$ ,

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

Now  $\mathbb{E}_f \left( (\psi_n^{-1} \|f - \theta_n^*\|_\infty)^{2\gamma} \right) \leq \psi_n^{-2\gamma} D_6 \left( Q^{2\gamma} + \mathbb{E}_f \|\theta_n^*\|_\infty^{2\gamma} \right)$ . 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  $\lim_{n \rightarrow \infty} \mathbb{E}_f \left[ w \left( \psi_n^{-1} \|f - \theta_n^*\|_\infty \right) I_{B_n^C} \right] = 0$ .

### 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

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

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

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

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

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

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  $X_1, \dots, X_n$ , we have

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

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  $\sigma^2$  in terms of  $n$ ,  $C_0$ ,  $L$ ,  $\beta$  and  $\mu_0$ , 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  $B_n$

$$\begin{aligned} \sum_{j=1}^n \frac{1}{nh} K \left( \frac{X_j - x_k}{h} \right) &\geq \mu(x_k) - \delta_n \geq \mu_0 - \delta_n, \\ \sum_{j=1}^n \frac{1}{nh} K^2 \left( \frac{X_j - x_k}{h} \right) &\leq \frac{\beta + 1}{2\beta + 1} \mu(x_k) + \delta_n. \end{aligned}$$

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  $n \rightarrow \infty$ . Because of the fact that  $\text{card}\{k\} \leq \delta_n^{-1} \psi_n^{-1/\beta}$ , 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  $\tilde{U}_n(x_k, f) = U_n(x_k, f)P^X(B_n)I_{B_n} - \mathbb{E}_f[U_n(x_k, f)P^X(B_n)I_{B_n}]$ . If  $B_n$  holds, we have  $\tilde{Z}_n(x_k, f) = \tilde{U}_n(x_k, f) - \mathbb{E}_f[U_n(x_k, f)I_{B_n}]P^X(B_n^C)$ . 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

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

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

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

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  $\tilde{U}_n(x_k, f)$  (note that here the family of random variables contains only one summand), we obtain

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

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

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

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

### 3.3.6 Proof of Lemma 2

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

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

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

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

Thus

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

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

### 3.3.7 Proof of Proposition 5

(i) The fact that  $y_j$  is conditionally gaussian with conditional mean  $\theta_j$  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  $X \in N_n$

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

Using that  $X \in N_n$  and replacing the expression for  $\sigma$  and  $h$  in terms of  $C_0$ ,  $n$ ,  $\beta$  and  $L$ , we obtain the inequality for  $v_j$ .

(ii) comes from the fact the functions  $f_j$  have disjoint supports and that the  $\xi_i$ 's are independent and independent of the  $X_i$ 's. (iii) is obtained by calculating the likelihood function of  $Y_1, \dots, Y_n$  conditionally on  $X$ , for  $X \in N_n$ .

## References

- [1] Donoho D., 1994. Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. Probability Theory and Related Fields. 99, 145-170.
- [2] Efromovich S., 1999. Nonparametric Curve Estimation: Methods, Theory and Applications, New York: Springer.



- [3] Ibragimov I.A., Has'minskii R.Z., 1981. *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, New-York.
- [4] Ibragimov I.A., Has'minskii R.Z., 1982. Bounds for the risks of nonparametric regression estimates. *Theory of Probability and its Applications*. 27, 84-99.
- [5] Korostelev A.P., 1993. Asymptotically minimax regression estimator in the uniform norm up to the exact constant. *Theory of Probability and its Applications*. 38, 737-743.
- [6] Korostelev A.P., Nussbaum M, 1999. The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*. 5, 1099-1118.
- [7] Lepski O.V., 1992. On problems of adaptive estimation in white noise. *Advances in Soviet Mathematics*. 12, 87-106.
- [8] Lepski O.V., Tsybakov A.B., 2000. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*. 117, 17-48.
- [9] Leonov S.L., 1997. On the solution of an optimal recovery problem and its applications in nonparametric regression. *Mathematical Methods of Statistics*. 4, 476-490.
- [10] Stone C.J., 1982. Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*. 13, 1040-1053.
- [11] Tsybakov A.B., 2003. *Introduction à l'estimation non-paramétrique*, Springer.