



**HAL**  
open science

# Rates of convergence for pointwise curve estimation with a degenerate design

Stéphane Gaïffas

► **To cite this version:**

Stéphane Gaïffas. Rates of convergence for pointwise curve estimation with a degenerate design. 2004.  
hal-00003086v1

**HAL Id: hal-00003086**

**<https://hal.science/hal-00003086v1>**

Preprint submitted on 15 Oct 2004 (v1), last revised 13 Jan 2006 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# RATES OF CONVERGENCE FOR POINTWISE CURVE ESTIMATION WITH A DEGENERATE DESIGN

STÉPHANE GAÏFFAS

*Laboratoire de Probabilités et Modèles Aléatoires, U.M.R. CNRS 7599  
and Université Paris 7, 175 rue du Chevaleret, 75013 Paris  
email: gaiffas@math.jussieu.fr*

ABSTRACT. The model of nonparametric regression with a random design is considered. We want to estimate the regression function at a point  $x_0$  where the density of the design is vanishing or exploding. Depending on assumptions on the local regularity of the regression function and on the local behaviour of the design, we find several minimax rates. These rates lie in a wide range, from slow rates of order  $\ell(n)$  where  $\ell(n)$  is slowly varying (for instance  $(\log n)^{-1}$ ) to fast rates of order  $n^{-1/2}\ell(n)$ . In particular, if the modulus of continuity at  $x_0$  of the regression function can be bounded from above by a regularly varying function of index  $s$ , and if the density of the design is regularly varying of index  $\beta$ , we prove that the minimax rate of convergence at  $x_0$  is of order  $n^{-s/(1+2s+\beta)}\ell(n)$ .

## 1. INTRODUCTION

1.1. **The model.** Suppose that we have  $n$  independent and identically distributed observations  $(X_i, Y_i) \in \mathbb{R} \times \mathbb{R}$  such that:

$$(1.1) \quad Y_i = f(X_i) + \xi_i,$$

where  $f : \mathbb{R} \rightarrow \mathbb{R}$  is the regression function and where the variables  $(\xi_i)_{i=1, \dots, n}$  (the *noise*) are centered Gaussian of variance  $\sigma^2$  and independent of  $\mathfrak{X}_n \triangleq \sigma(X_1, \dots, X_n)$ , the  $\sigma$ -algebra generated by the variables  $(X_i)_{i=1, \dots, n}$ . The variable  $Y_i$  is then a noisy observation of the function  $f$  at the random point  $X_i$ . Let  $x_0 \in \mathbb{R}$  be the estimation point. The random variables  $(X_i)_{i=1, \dots, n}$  (the *design*) are distributed with respect to a density  $\mu$ . Take for instance the density

$$\mu(x) = \frac{\beta + 1}{x_0^{\beta+1} + (1 - x_0)^{\beta+1}} |x - x_0|^\beta \mathbf{1}_{[0,1]}(x),$$

for  $x_0 \in [0, 1]$  and  $\beta > -1$ . When  $\beta > 0$  this density models a lack of information at the point  $x_0$  quantified by the parameter  $\beta$ . Conversely, when  $-1 < \beta < 0$ ,  $\mu$  models an exploding quantity of information at  $x_0$  quantified by  $\beta$ .

1.2. **Motivations.** The pointwise estimation of a regression function is a well-known problem which has been intensively studied by many authors. The first authors who computed the minimax rate over a nonparametric class of Hölderian functions are Ibragimov and Hasminski (1981) and Stone (1980). Over a class of Hölderian functions of smoothness parameter  $s$ , the local polynomial estimator converges with the rate  $n^{-s/(1+2s)}$  and this rate is

---

*Date:* October 15, 2004.

*2000 Mathematics Subject Classification.* 62G05, 62G08, 62K99.

*Key words and phrases.* degenerate design, minimax, nonparametric regression, random design.

I wish to thank my adviser Marc Hoffmann for helpful suggestions and encouragements.

optimal in the minimax sense. Many authors worked on related problems: see for instance Korostelev and Tsybakov (1993), Nemirovski (2000), Tsybakov (2003). Nevertheless, these results require that the density of the design is non-vanishing and finite at the point of estimation. This assumption roughly means that the information is spatially *homogeneous*. The next logical step is then to look for the minimax risk at a point where the density of the design  $\mu$  is vanishing or exploding. Such a result would improve the understanding of the case of very *inhomogeneous* distribution of the observations. Hall *et al.* (1997) answer this question in the case of a Hölder type smoothness of order 2: if  $\mu(x) \sim x^\beta$  near 0, where  $\beta > 0$ , the authors show that a local linear procedure converges with the rate  $n^{-4/(5+\beta)}$  when estimating  $f$  at 0. The optimality of this rate is also proved. Guerre (1999) considers in a more general setup for the design the case of a Lipschitz regression function and in particular extends the result of Hall *et al.* for  $\beta > -1$ . We intend here to develop the estimation of the regression function when the design is degenerate in a systematical way.

**1.3. Results.** The rates of convergence are given in the minimax sense. Let us define the pointwise minimax risk over the class  $\Sigma$  (see 3.2 for a definition):

$$(1.2) \quad \mathcal{R}_{n,p}(\Sigma, \mu, x_0) \triangleq \left( \inf_{T_n} \sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \{|T_n - f(x_0)|^p\} \right)^{1/p},$$

where  $\inf_{T_n}$  stands for any estimator  $T_n$  based on the observations (1.1),  $x_0$  is the point of estimation and  $p$  is some positive number. The expectation  $\mathbb{E}_{f,\mu}^n$  in (1.2) is taken with respect to the joint probability  $\mathbb{P}_{f,\mu}^n$  of pairs of random variables  $(X_i, Y_i)_{i=1,\dots,n}$ . In this paper we prove that the following minimax rates hold:

$$r_n \asymp n^{-s/(1+2s+\beta)} \ell(n) \text{ as } n \rightarrow +\infty,$$

where  $a_n \asymp b_n$  means  $0 < \liminf_n a_n/b_n \leq \limsup_n a_n/b_n < +\infty$ , and where  $s > 0$  is the local (around  $x_0$ ) smoothness parameter of the regression function, and  $\beta$  is a parameter describing the local behaviour of the design density. More precisely, if the modulus of continuity of the regression function  $f$  at  $x_0$  is bounded from above by a regularly varying function  $\omega$  at 0 of index  $s$  (if  $\omega(h) = rh^s$  for instance, we find a classical Hölder ball of radius  $r$ ) and if the design density  $\mu$  has the form  $\nu(|\cdot - x_0|)$  close to  $x_0$  where  $\nu$  is regularly varying of index  $\beta \geq -1$ , the minimax rate  $r_n$  of estimation at  $x_0$  is of order:

$$n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(n),$$

where  $\ell_{\omega,\nu}$  is a slowly varying function depending on  $\omega$  and  $\nu$ . When  $\beta = -1$  it is noteworthy that

$$r_n \asymp n^{-1/2} \ell_{\omega,\nu}(n),$$

which is barely a minimax parametric rate of estimation up to the slow term  $\ell$ . We consider also the framework of  $\Gamma$ -variation for the density of the design, including densities behaving like  $\exp(-1/|x - x_0|^\alpha)$  for  $\alpha > 0$  at  $x_0$  (which is an example of function vanishing at  $x_0$  faster than any power function). We show in this  $\Gamma$ -variation framework that the minimax rate is slow, and in this latter example

$$r_n \asymp (\log n)^{-s/\alpha}.$$

When the design is regularly varying or  $\Gamma$ -varying, we prove that

$$r_n = \omega(h_n),$$

where  $h_n$  (the *bandwidth*) is the smallest solution of the *bias-variance* equation:

$$\omega(h) = \frac{\sigma}{\sqrt{nF_\nu(h)}},$$

where  $\sigma$  is the noise level ( $\sigma^2$  is the variance of the  $\xi_i$ ) and where  $F_\nu(h) = \int_0^h \nu(t)dt$ .

**1.4. Adaptation.** Forthcoming results concern adaptation in regularity and design. The procedure considered in this paper heavily depends (via the *bandwidth*  $h_n$ ) on the smoothness parameter of the regression function and on the behaviour of the density of the design at the point of estimation. An extension of this result is then to look for a procedure able to estimate the regression function without an a priori knowledge on the regularity of the regression function nor the density of the design. On design adaptation, see Guerre (1999), Guerre (2000) and on pointwise regularity adaptation see Cai and Brown (1998), Lepski (1990), Lepski *et al.* (1997), Lepski and Spokoiny (1997), Nemirovski (2000), Spokoiny (1998) among many others.

**1.5. Organisation of the paper.** We introduce in section 2 the local polynomial estimation procedure, we define our estimator in definition 2.1. We give a bias-variance decomposition in proposition 2.3, conditional on the design. In section 3 we assume the design is regularly varying, and we give upper and lower bounds in theorems 3.6 and 3.7. In section 4 we assume the design is  $\Gamma$ -varying. The upper and lower bounds are stated in theorems 4.7 and 4.8. Section 5 is devoted to the proofs, and the sections A and B recall technical results of the Karamata theory (regular variation) and the de Haan theory ( $\Pi$ -variation,  $\Gamma$ -variation).

## 2. LOCAL POLYNOMIAL ESTIMATION

**2.1. Introduction.** The local polynomial estimator is well-known and has been intensively studied (see Stone (1980), Fan and Gijbels (1996), Spokoiny (1998), Tsybakov (2003), among others). If  $f$  is a smooth function at  $x_0$ , it is close to its Taylor polynomial at  $x_0$ : a function  $f \in C^k(x_0)$  (the space of  $k$  times differentiable functions at  $x_0$  with a continuous  $k$ -th derivative) is such that for any  $x$  close to  $x_0$ :

$$(2.1) \quad f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \dots + \frac{f^{(k)}(x_0)(x - x_0)^k}{k!}.$$

Let  $h > 0$  (the *bandwidth*) and  $k \in \mathbb{N}$ . We define  $\phi_{j,h}(x) \triangleq \left(\frac{x-x_0}{h}\right)^j$  and the space

$$V_{k,h} \triangleq \text{Span}\{(\phi_{j,h})_{j=0,\dots,k}\}.$$

For some fixed non-negative function  $K$  (the *kernel*) we define the weighted pseudo-scalar product:

$$(2.2) \quad \langle f, g \rangle_{h,K} \triangleq \sum_{i=1}^n f(X_i)g(X_i)K\left(\frac{X_i - x_0}{h}\right),$$

and  $\|\cdot\|_{h,K}$  the corresponding pseudo-norm ( $K \geq 0$ ). In view of (2.1) it is natural to consider the estimator defined as the closest polynomial of order  $k$  to the observations  $(Y_i)_{i=1,\dots,n}$  in the least square sense, that is:

$$(2.3) \quad \hat{f}_h = \arg \min_{g \in V_{k,h}} \|g - Y\|_{h,K}^2.$$

Then  $\hat{f}_h(x_0)$  is the *local polynomial estimator* of  $f$  at  $x_0$ . A necessary condition for  $\hat{f}_h$  to be the minimizer of (2.3) is to be solution of the linear problem:

$$(2.4) \quad \text{find } \hat{f} \in V_{k,h} \text{ such that } \forall \phi \in V_{k,h}, \quad \langle \hat{f}, \phi \rangle_{h,K} = \langle Y, \phi \rangle_{h,K}.$$

The estimator  $\widehat{f}_h$  is given by:

$$(2.5) \quad \widehat{f}_h = P_{\widehat{\theta}_h},$$

where

$$(2.6) \quad P_{\theta} = \theta_0 \phi_{0,h} + \theta_1 \phi_{1,h} + \dots + \theta_k \phi_{k,h},$$

with  $\widehat{\theta}_h$  the solution, whenever it makes sense, of the linear system:

$$(2.7) \quad \mathbf{X}_h^K \theta = \mathbf{Y}_h^K,$$

where  $\mathbf{X}_h^K$  is the symmetrical matrix with entries, for  $0 \leq j, l \leq k$ :

$$(2.8) \quad (\mathbf{X}_h^K)_{j,l} = \langle \phi_{j,h}, \phi_{l,h} \rangle_{h,K},$$

and  $\mathbf{Y}_h^K$  is the vector defined by:

$$\mathbf{Y}_h^K = (\langle Y, \phi_{j,h} \rangle_{h,K}; 0 \leq j \leq k).$$

We assume that the kernel  $K$  satisfies the following assumptions:

**Assumption K.**  $K$  is a non-negative function such that:

- $\text{Supp } K \subset [-1, 1]$ ,
- $K$  is symmetrical,
- $K_{\infty} \triangleq \sup_x K(x) \leq 1$ ,
- There is some  $\rho > 0$  and  $\kappa > 0$  such that  $\forall x, y \in \text{Supp } K, |K(x) - K(y)| \leq \rho|x - y|^{\kappa}$ .

*Remark.* The conditions on the kernel  $K$  are satisfied by all the classical kernels used in nonparametric curve smoothing. The first assumption is used to make the kernel  $K$  localise the information around the point of estimation  $x_0$  (see (2.2)). The last one is technical. The two other ones are used for the sake of simplicity, since we only really need the kernel to be bounded from above.

Let us define:

$$N_{n,h} = \#\{X_i \text{ such that } X_i \in [x_0 - h, x_0 + h]\},$$

the number of design points falling in the interval  $[x_0 - h, x_0 + h]$ , and we define the random matrix (measurable with respect to  $\mathfrak{X}_n$ )

$$\mathcal{X}_h^K \triangleq N_{n,h}^{-1} \mathbf{X}_h^K.$$

The matrix  $\mathcal{X}_h^K$  is a "renormalisation" of the matrix  $\mathbf{X}_h^K$ . We will see that this matrix is asymptotically non-degenerate with a large probability in the context of section 3 (see proposition 5.2). Actually our estimator is a bit different from the classical local polynomial estimator (defined as the solution of the linear system (2.7)) that can be found in the literature. For technical reasons, we need to introduce a "correction" term in the matrix  $\mathbf{X}_h^K$ . For a fixed bandwidth  $h > 0$ , our estimator is defined as follows:

**Definition 2.1.** Given some  $h > 0$ , we consider  $\widehat{f}_h$  defined by (2.5) with  $\widehat{\theta}_h$  the solution, when it makes sense (if  $N_{n,h} = 0$  we take  $\widehat{f}_h = 0$ ), of the linear system:

$$(2.9) \quad \widetilde{\mathbf{X}}_h^K \theta = \mathbf{Y}_h^K,$$

where:

$$\widetilde{\mathbf{X}}_h^K \triangleq \mathbf{X}_h^K + N_{n,h}^{1/2} \mathbf{I}_{k+1} \mathbf{1}_{\lambda(\mathbf{X}_h^K) \leq N_{n,h}^{1/2}},$$

with  $\lambda(M)$  standing for the smallest eigenvalue of a matrix  $M$ , and  $\mathbf{I}_{k+1}$  denoting the identity matrix in  $\mathbb{R}^{k+1}$ .

*Remark.* One can understand the definition of the matrix  $\tilde{\mathbf{X}}_h^K$  as follows: in the "good" case, that is when the matrix  $\mathcal{X}_h^K$  is non-degenerate (the smallest eigenvalue of  $\mathcal{X}_h^K$  is not too small:  $\lambda(\mathcal{X}_h^K) > N_{n,h}^{-1/2}$ ), we solve the system (2.7), while in the "bad" case we still have a control on the smallest eigenvalue of  $\tilde{\mathbf{X}}_h^K$  (that we need in the proof of the upper bound), since  $\lambda(\tilde{\mathbf{X}}_h^K) \geq N_{n,h}^{1/2}$ .

**2.2. Bias-variance equilibrium.** A main result on the local polynomial estimator is the bias-variance decomposition. This is a classical result, many times presented in different forms: see Cleveland (1979), Goldenshluger and Nemirovski (1997), Korostelev and Tsybakov (1993), Spokoiny (1998), Stone (1977), Tsybakov (1986) and Tsybakov (2003). The version by Spokoiny (1998) is close to the one presented here. Mainly, the differences are linked with the fact that the design is random and that we consider a slightly modified version of the local polynomial estimator (see definition 2.1). Let us first define the modulus of continuity at  $x_0$ :

**Definition 2.2** (Modulus of continuity). Let  $k \in \mathbb{N}$ . The modulus of continuity of a continuous function  $f$  at  $x_0$  is defined by:

$$(2.10) \quad \omega_{f,k}(x_0, h) = \inf_{P \in \mathcal{P}_k} \sup_{|x-x_0| \leq h} |f(x) - P(x-x_0)|,$$

where  $\mathcal{P}_k$  is the set of real polynomials of order  $k$ .

*Remark.* It is easy to see when  $f$  belongs to a Hölder ball of regularity  $s$  and radius  $r$  that  $\omega_{f,k}(x_0, h) \leq rh^s/(k!)$  where  $k \triangleq \lfloor s \rfloor$  is the largest integer smaller than  $s$ .

We introduce the event

$$(2.11) \quad \Omega_h^K \triangleq \{X_1, \dots, X_n \text{ are such that } \lambda(\mathcal{X}_h^K) > N_{n,h}^{-1/2} \text{ and } N_{n,h} > 0\}.$$

Note that on  $\Omega_h^K$  the matrix  $\mathcal{X}_h^K$  is invertible.

**Proposition 2.3** (Bias-variance decomposition). *Under assumption K, we have on the event  $\Omega_h^K$  the following decomposition:*

$$(2.12) \quad |\hat{f}_h(x_0) - f(x_0)| \leq \lambda^{-1}(\mathcal{X}_h^K) \sqrt{k+1} K_\infty (\omega_{f,k}(x_0, h) + \sigma N_{n,h}^{-1/2} |\gamma_h|),$$

where  $\gamma_h$  is, conditional on  $\mathfrak{X}_n$ , a centered Gaussian random variable s.t.  $\mathbb{E}_{f,\mu}^n \{\gamma_h^2 | \mathfrak{X}_n\} \leq 1$ .

*Remark.* Inequality (2.12) in proposition 2.3 holds conditional on the design, on the event  $\Omega_h^K$ . We will see this event has a large probability in the regular variation framework.

Now the problem is, like with any linear estimation procedure, to answer the following question: **how to choose the bandwidth  $h$** ? A preliminary work is to study the behaviour of the variance term  $\sigma^2 N_{n,h}^{-1/2}$  in the decomposition (2.12). In order to do so, we need some assumptions on the local behaviour of the law of the design close to the point of estimation  $x_0$ . We assume in all the following:

**Assumption M.** There is some fixed neighbourhood  $W$  of  $x_0$  and some continuous function  $\nu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that:

$$(2.13) \quad \forall x \in W, \quad \mu(x) = \nu(|x - x_0|).$$

In particular, we assume the law of the design is symmetrical close to  $x_0$ . The following result makes a link between the behaviour of the counting process  $N_{n,h}$  (that appears in the variance term of the decomposition (2.12)) and the behaviour of  $\mu$  close to  $x_0$ . Actually,

the regular variation property (defined in section 3) naturally appears under assumptions on the asymptotic behaviour of  $N_{n,h}$ . This result will motivate assumption R below, that is the choice of the regular variation framework for the design. Let us denote by  $\mathbb{P}_\mu^n$  the joint probability of the variables  $(X_i)_{i=1,\dots,n}$ .

**Proposition 2.4.** *Under assumption M, with  $\nu$  monotone, the following properties are equivalent:*

- (1)  $\nu$  is regularly varying of index  $\beta \geq -1$ ,
- (2) There exist sequences of positive numbers  $(\lambda_n)_{n \geq 0}$  and  $(\gamma_n)_{n \geq 0}$ , such that  $\lim_n \gamma_n = 0$ ,  $\liminf_n n\lambda_n^{-1} > 0$ ,  $\gamma_{n+1} \sim \gamma_n$  as  $n \rightarrow +\infty$  and a continuous function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that for any  $C > 0$ :

$$\mathbb{E}_\mu^n \{N_{n,C\gamma_n}\} \sim \phi(C)\lambda_n \text{ as } n \rightarrow +\infty,$$

- (3) There exist  $(\lambda_n)_{n \geq 0}$ ,  $(\gamma_n)_{n \geq 0}$  and  $\phi$  as previously such that for any  $C > 0$  and  $\varepsilon > 0$ :

$$\lim_{n \rightarrow +\infty} \frac{n}{\lambda_n} \mathbb{P}_\mu^n \left\{ \left| \frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1 \right| > \varepsilon \right\} = 0.$$

The proof is delayed until section 5. It is a consequence of the sequence characterisation of regular variation (see theorem A.10 for the details).

### 3. REGULARLY VARYING DESIGN

**3.1. Introduction.** The definition and the main properties of regularly varying functions are due to Karamata (1930). Main references on regular variation are Bingham *et al.* (1989), Geluk and de Haan (1987), Resnick (1987) and Senata (1976).

**Definition 3.1** (Regular variation). A continuous function  $\nu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is regularly varying at 0 if there is a function  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  and a real number  $\beta \in \mathbb{R}$  such that:

$$(3.1) \quad \forall y > 0, \quad \lim_{h \rightarrow 0^+} \nu(yh)/\nu(h) = y^\beta.$$

We denote by  $\text{RV}_0(\beta)$  the space of functions satisfying (3.1). A function in  $\text{RV}_0(0)$  is said *slowly varying*.

**Assumption R.** Assumption M holds with  $\nu \in \text{RV}_0(\beta)$  for  $\beta \geq -1$ . We denote by  $\ell_\nu(h) \triangleq h^{-\beta}\nu(h)$  the slow term of  $\nu$ .

*Remark.* Typical examples of regularly varying function (of index  $\beta$ ) are  $x^\beta$ ,  $x^\beta(\log(1/x))^\gamma$  for  $\gamma \in \mathbb{R}$ , and more generally any power function times a log or compositions of log to some power. For other examples, see in the references cited above.

The aim of the following part is to compute the minimax risk for the problem of estimation at  $x_0$  when the design is regularly varying at this point in the sense of assumption R. First we prove upper bounds using the local polynomial estimator, and secondly we prove the corresponding lower bounds, so the minimax rates will follow.

### 3.2. The upper bound.

**Definition 3.2.** For  $s > 0$ , a sequence of positive numbers  $(\delta_n)_n$  going to 0 and  $\omega \in \text{RV}_0(s)$ , we define

$$\mathcal{F}_{\delta_n}(x_0, \omega) \triangleq \{f : [0, 1] \rightarrow \mathbb{R} \text{ such that } \forall h \leq \delta_n, \quad \omega_{f,k}(x_0, h) \leq \omega(h)\},$$

where  $k = \lfloor s \rfloor$  (the largest integer smaller than  $s$ ). We define  $\ell_\omega(h) \triangleq \omega(h)h^{-s}$ , the slow term of  $\omega$ . For a sequence of positive numbers  $(\alpha_n)_n$  going to  $+\infty$ , we define

$$\mathcal{U}(x_0, \delta_n, \alpha_n) \triangleq \left\{ f : [0, 1] \rightarrow \mathbb{R} \text{ such that } \sup_{x \in [x_0 - \delta_n, x_0 + \delta_n]} |f(x)| \leq \alpha_n \right\}.$$

Finally, we define

$$\Sigma_{\delta_n, \alpha_n}(x_0, \omega) \triangleq \mathcal{F}_{\delta_n}(x_0, \omega) \cap \mathcal{U}(x_0, \delta_n, \alpha_n).$$

*Remark.* In a classical Hölder ball, we estimate the modulus of continuity at  $x_0$  by some power function times a constant (the radius). The above class generalizes the Hölder regularity, since it estimates the modulus of continuity by a power function times a slow factor. Thus we find back a classical Hölder balls of radius  $r$  by choosing the slowly varying function  $\ell_\omega$  constant and equal to  $r$ .

*Remark.* Since we are interested in the pointwise estimation of the regression function  $f$ , we only need a regularity assumption in some neighbourhood of  $x_0$ . In definition 3.2 we note that the class of functions depends on the number of observations  $n$ , and that we only ask for a control on the modulus of continuity in a decreasing interval  $[0, \delta_n]$  as  $n$  increases. We will see that a natural choice for  $\delta_n$  is the optimal bandwidth  $h_n$ , defined below.

**Proposition 3.3.** *Under assumption R, one has, for any positive sequence  $\gamma_n$  going to 0:*

$$\forall \varepsilon > 0, \quad \mathbb{P}_\mu^n \left\{ \left| \frac{N_{n, \gamma_n}}{2nF_\nu(\gamma_n)} - 1 \right| > \varepsilon \right\} \leq 2 \exp(-\varepsilon^2 C_1(\beta) n F_\nu(\gamma_n)),$$

where  $C_1(\beta) = (3/(\beta + 1) + 4/3)^{-1}/2$  when  $\beta > -1$  and  $C_1(\beta) = 64/13$  when  $\beta = -1$ .

Then, when  $nF_\nu(\gamma_n) \rightarrow +\infty$  as  $n \rightarrow +\infty$ , we have:

$$N_{n, \gamma_n} \underset{\mathbb{P}_\mu^n}{\sim} 2nF_\nu(\gamma_n) \text{ as } n \rightarrow +\infty,$$

where  $X_n \underset{\mathbb{P}}{\sim} Y_n$  means  $\lim_n \mathbb{P}\{|X_n/Y_n - 1| > \varepsilon\} = 0$  for any  $\varepsilon > 0$ . The equation (2.12) and the proposition 3.3 together entail, when  $f \in \mathcal{F}_{\delta_n}(x_0, \omega)$ , (the term  $\lambda(\mathcal{X}_h^K)$  is of the order of a positive constant when  $h \rightarrow 0^+$ , see lemma 5.2) that a natural choice for the optimal bandwidth  $h = h_n$  is the following:

**Definition 3.4.** Let  $h_n$  be defined as the smallest solution (well defined in view of proposition 3.5 below) of the equation:

$$(3.2) \quad \omega(h) = \frac{\sigma}{\sqrt{nF_\nu(h)}},$$

where  $F_\nu(h) = \int_0^h \nu(t) dt$ .

The following proposition gives the existence, the uniqueness and the form of  $h_n$ :

**Proposition 3.5.** *Under assumption R and if  $\omega \in \text{RV}_0(s)$  for some  $s > 0$ , there is a solution to equation 3.2 for  $n$  large enough. We have  $\lim_{n \rightarrow +\infty} h_n = 0$  and:*

$$(3.3) \quad h_n \sim \sigma^{2/(1+2s+\beta)} n^{-1/(1+2s+\beta)} \ell_{\omega, \nu}(1/n) \text{ as } n \rightarrow +\infty,$$

where  $\ell_{\omega, \nu}$  is a slowly varying function. When  $\omega(h) = rh^s$  (Hölder regularity) for some  $r > 0$ , we have more precisely:

$$(3.4) \quad h_n \sim (\sigma/r)^{2/(1+2s+\beta)} n^{-1/(1+2s+\beta)} \ell_{s, \nu}(1/n) \text{ as } n \rightarrow +\infty,$$



where  $\ell_{s,\nu}$  is again slowly varying. Moreover, if  $h_{n,1}$  and  $h_{n,2}$  are both solutions of (3.2), then  $h_{n,1} \sim h_{n,2}$ . It is noteworthy that these results still hold when  $\beta = -1$ . In this case, (3.3) becomes:

$$h_n \sim \sigma^{1/s} n^{-1/(2s)} \ell_{\omega,\nu}(1/n) \text{ as } n \rightarrow +\infty.$$

Then we plug-in the bandwidth  $h_n$  in the estimator  $\widehat{f}_h$  (see definition 2.1):

$$(3.5) \quad \widehat{f}_n(x_0) \triangleq \widehat{f}_{h_n}(x_0).$$

The next result tells that the rate

$$(3.6) \quad r_n \triangleq \omega(h_n),$$

for  $h_n$  given by definition 3.4 is a rate of convergence for the estimator  $\widehat{f}_n(x_0)$  over the class  $\Sigma_{h_n, \alpha_n}(x_0, \omega)$ , where  $\alpha_n = n^\gamma$ , for any  $\gamma > 0$ . Note that  $\gamma$  needs not to be known by the statistician. We need to introduce some notation. Let  $C(p) \triangleq \sqrt{2/\pi} \int_{\mathbb{R}^+} (1+t)^p \exp(-t^2/2) dt$ . We define  $\lambda_\beta^K > 0$  as the smallest eigenvalue of the symmetrical and positive matrix with entries, for  $0 \leq j, l \leq k$ :

$$(3.7) \quad (\mathcal{X}_\beta^K)_{j,l} = \frac{\beta+1}{2} (1 + (-1)^{j+l}) \int_0^1 y^{j+l+\beta} K(y) dy.$$

We choose the sequence  $\delta_n = h_n$  in definition 3.2, since we only need to control the bias term  $\omega_{f,k}(x_0, h)$  (see proposition 2.3) for  $h$  in the interval  $[0, h_n]$ .

**Theorem 3.6.** *Let us assume  $K$  and  $R$  with  $\beta > -1$ . We assume there exists  $\gamma > 0$  such that  $\alpha_n = n^\gamma$ . For any  $p > 0$ , the following inequality holds:*

$$(3.8) \quad \sup_{f \in \Sigma_{h_n, \alpha_n}(x_0, \omega)} \mathbb{E}_{f,\mu}^n \{r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p\} \leq C(p) (\lambda_\beta^K)^{-p} K_\infty^p (k+1)^{p/2} + o_n(1),$$

where the rate  $r_n$  defined by (3.6) satisfies:

$$r_n \sim \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(1/n) \text{ as } n \rightarrow +\infty,$$

where  $\ell_{\omega,\nu}$  is a slowly varying function. When assumption  $R$  holds with  $\beta = -1$ , and if  $K(0) > 0$ , we have for any  $0 < s \leq 1$ :

$$(3.9) \quad \sup_{f \in \Sigma_{h_n, \alpha_n}(x_0, \omega)} \mathbb{E}_{f,\mu}^n \{r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p\} \leq C(p) K(0)^{-p} K_\infty^p + o_n(1),$$

where

$$r_n \sim \sigma n^{-1/2} \ell_{\omega,\nu}(1/n) \text{ as } n \rightarrow +\infty.$$

*Remark.* When  $\ell_\omega$  is constant and equal to  $r > 0$  ( $f$  is in an Hölder ball of radius  $r$ ) we have:

$$r_n \sim \sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{s,\nu}(1/n) \text{ as } n \rightarrow +\infty.$$

*Remark.* It is important to notice when  $\beta = -1$  that the upper bound (3.9) only holds for small regularities  $0 < s \leq 1$ , for technical reasons. In this case, it is clear in view of (2.4) (here  $k = 0$ ) that the local polynomial estimator is a Nadaraya-Watson estimator, defined as:

$$\widehat{f}_n(x_0) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x_0}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_0}{h_n}\right)}.$$

### 3.3. The lower bound.

**Theorem 3.7.** *Under assumption R one has for any  $p > 0$  and  $n$  large enough:*

$$(3.10) \quad \mathcal{R}_{n,p}(\Sigma_{h_n, \alpha_n}(x_0, \omega), \mu, x_0) \geq C_4 r_n,$$

where  $r_n$  is given by (3.6) and  $C_4$  is some constant depending on  $s$  and  $p$  only.

Then theorems 3.6 and 3.7 together entail that  $r_n$  is the minimax rate of estimation at the point  $x_0$  over the class  $\Sigma_{h_n, \alpha_n}(x_0, \omega)$ .

### 3.4. Examples.

**First example.** The simplest example is the case of a non-degenerate design ( $0 < \mu(x_0) < +\infty$ ) with a modulus of continuity of the regression function bounded from above by  $\omega(h) = rh^s$ . This is the common case found in the literature. In this case, the design is actually slowly varying ( $\beta = 0$ , with the slow term constant and equal to the limit at  $x_0$ ), and solving (3.2) leads to the classical minimax rate:

$$r_n \asymp \sigma^{2s/(1+2s)} r^{1/(1+2s)} n^{-s/(1+2s)},$$

where  $\asymp$  denotes here and in all this section the equality in order, up to constants depending on  $s$  and  $\beta$ , but not on  $r$  nor  $\sigma$ .

**Second example.** We choose  $\alpha, \gamma \in \mathbb{R}$  and  $r > 0, s > 0, \beta > -1$ . We consider  $\nu$  such that  $F_\nu(h) = h^{\beta+1}(\log(1/h))^\alpha$  and  $\omega(h) = rh^s(\log(1/h))^\gamma$ . Then  $h_n$  satisfies, in view of lemma 5.5 (see the proof of proposition 3.5 and the proof of lemma 5.5 for the details):

$$h_n \sim (1 + 2s + \beta)^{(\alpha+2\gamma)/(1+2s+\beta)} (\sigma/r)^{2/(1+2s+\beta)} (n(\log n)^{\alpha+2\gamma})^{-1/(1+2s+\beta)},$$

and then we find the minimax rate:

$$r_n \asymp \sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)} (n(\log n)^{\alpha-\gamma(1+\beta)/s})^{-s/(1+2s+\beta)}.$$

*Remark.* We note this rate has the form given by theorem 3.6 with the slow factor  $\ell_{\omega, \nu}(h) = (\log(1/h))^{\gamma(\beta+1)-s\alpha}/(1+2s+\beta)$ . When  $\gamma(1+\beta) - s\alpha = 0$  there is no slow term in the minimax rate, although there are slow terms in  $\nu$  and  $\omega$ . Again, if  $\beta = 0$  and  $\gamma = s\alpha$ , we find the classical minimax rate of the first example, although the terms  $\nu$  and  $\omega$  do not have the classical forms.

**Third example.** We consider the case  $\beta = -1$ . We choose  $\alpha > 1$ , thus  $\nu(h) = (\alpha - 1)h^{-1}(\log(1/h))^{-\alpha}$  is integrable at 0. We have  $F_\nu(h) = (\log(1/h))^{1-\alpha}$ . We take the same  $\omega$  as in the second example, with  $0 < s \leq 1$  (when  $\beta = -1$ , theorem 3.6 holds only for small regularities). Then we have:

$$h_n \sim (2s)^{(2\gamma+1-\alpha)/(2s)} (\sigma/r)^{1/s} n^{-1/(2s)} (\log n)^{(\alpha-1-2\gamma)/(2s)},$$

and then:

$$r_n \asymp \sigma n^{-1/2} (\log n)^{(\alpha-1)/2}.$$

*Remark.* This rate is barely the parametric rate of estimation (up to the slow log factor). This is due to the fact that since the design is very "exploding", we have a lot of information at the point of estimation  $x_0$ , thus we can estimate  $f(x_0)$  very fast. Also, we note the parameters of regularity of the regression function (the radius term  $r$  and  $\gamma$ ) have disappeared from the minimax asymptotic rate.

4.  $\Gamma$ -VARYING DESIGN

**4.1. Introduction.** The framework of regular variation includes any design density behaving close to the point of estimation as a polynomial times a slow term. It does not include for instance a design with a behaviour similar to  $\exp(-1/|x - x_0|)$ , prolonged at  $x_0$  by 0. This function converges to 0 at  $x_0$  faster than any power function. Such a choice for the density of the design can model a very big lack of information at  $x_0$ , thus it is natural to look for a framework including such behaviours, and for the corresponding pointwise minimax risk. In fact such a function belongs to the following class, introduced by de Haan (1970):

**Definition 4.1** ( $\Gamma$ -variation). We denote by  $\Gamma V_0$ , the class of all non-decreasing and continuous functions  $\nu : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  such that there exists a continuous function  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  with:

$$(4.1) \quad \forall y \in \mathbb{R}, \quad \lim_{h \rightarrow 0^+} \nu(h + y\rho(h))/\nu(h) = \exp(y).$$

We denote by  $\Gamma V_0(\rho)$  the class of all such functions. The function  $\rho$  is called the *auxiliary* function of  $\nu$ .

We delay the technical properties of this class of functions in section B. Here we assume the following:

**Assumption G.** The variables  $(X_i)_{i=1\dots n}$  satisfy assumption M with a function  $\nu \in \Gamma V_0$ .

*Remark.* A density behaving like  $\exp(-1/|x - x_0|)$  close to  $x_0$  satisfies (2.13) with  $\nu(h) = \exp(-1/h)$ , where  $\nu$  belongs to  $\Gamma V_0(\rho)$  with the auxiliary function  $\rho(h) = h^2$ .

**4.2. The upper bound.**

**Definition 4.2.** For  $0 < s \leq 1$ , a sequence of positive numbers  $(\delta_n)_n$  going to 0 and  $\omega \in \text{RV}_0(s)$ , we define

$$\mathcal{F}_{\delta_n}(x_0, \omega) \triangleq \left\{ f : [0, 1] \rightarrow \mathbb{R} \text{ such that } \forall h \leq \delta_n, \quad \sup_{|x-x_0| \leq h} |f(x) - f(x_0)| \leq \omega(h) \right\},$$

and as previously we denote by  $\ell_\omega$  the slow term of  $\omega$ . For a sequence of positive numbers  $(\alpha_n)_n$  going to  $+\infty$  we define

$$\Sigma_{\delta_n, \alpha_n}(x_0, \omega) \triangleq \mathcal{F}_{\delta_n}(x_0, \omega) \cap \mathcal{U}(x_0, \delta_n, \alpha_n),$$

where the class  $\mathcal{U}$  is given in definition 3.2.

To prove the upper bounds, we can consider two Nadaraya-Watson type estimators. Let  $K$  be a kernel satisfying assumption K, and  $h > 0$  a bandwidth. Let  $(\eta_n)_{n \geq 0}$  be a sequence of positives numbers converging to 0, to be specified above. The first estimator is the regressogram:

$$(4.2) \quad \hat{f}_n(x_0, h) \triangleq \frac{\sum_{i=1}^n Y_i \mathbf{1}_{|X_i - x_0| \leq h}}{N_{n,h} \vee \eta_n},$$

where  $N_{n,h} = \sum_{i=1}^n \mathbf{1}_{|X_i - x_0| \leq h}$ . The second estimator is defined by:

$$(4.3) \quad \tilde{f}_n(x_0, h) \triangleq \frac{\sum_{i=1}^n Y_i K_{\rho,h}(X_i - x_0)}{\bar{K}_{n,\rho,h} \vee \eta_n},$$

where

$$(4.4) \quad K_{\rho,h}(x) \triangleq K\left(\frac{x-h}{\rho(h)}\right) + K\left(\frac{x+h}{\rho(h)}\right),$$

and

$$\overline{K}_{n,\rho,h} \triangleq \sum_{i=1}^n K_{\rho,h}(X_i - x_0).$$

Since  $\text{Supp } K \subset [-1, 1]$ , the estimator  $\tilde{f}_n(x_0, h)$  makes a local average of the observations  $Y_i$  such that  $X_i \in [x_0 - h - \rho(h), x_0 - h + \rho(h)] \cup [x_0 + h - \rho(h), x_0 + h + \rho(h)]$ , which is an interval not containing the point of estimation  $x_0$ . In spite of this,  $\tilde{f}_n(x_0, h)$  (for a good choice of the bandwidth) is minimax (see theorems 4.7, 4.8). If  $\mu$  satisfy assumption G, there is barely no information at  $x_0$ , so the procedure actually "catches" the information "far" from  $x_0$ .

**Proposition 4.3.** *Under assumption K, if  $f \in \mathcal{F}_{\delta_n}(x_0, \omega)$  and if  $0 < h \leq \delta_n$  is small enough and such that  $\overline{K}_{n,\rho,h} > \eta_n$  and  $N_{n,h} > \eta_n$ , we have:*

$$(4.5) \quad |\hat{f}_n(x_0, h) - f(x_0)| \leq \omega(h) + \sigma N_{n,h}^{-1/2} |\gamma_{h,1}|,$$

$$(4.6) \quad |\tilde{f}_n(x_0, h) - f(x_0)| \leq 2\omega(h) + \sigma \overline{K}_{n,\rho,h}^{-1/2} |\gamma_{h,2}|,$$

where  $(\gamma_{h,j})_{j=1,2}$  are, conditional on  $\mathfrak{X}_n$ , centered Gaussian variables s.t.  $\mathbb{E}_{f,\mu}^n \{\gamma_{h,j}^2 | \mathfrak{X}_n\} \leq 1$ .

As in the regular variation framework, the problem of the choice of an optimal bandwidth arises.

**Proposition 4.4.** *Under assumptions G, K one has for any sequence  $(\gamma_n)_{n \geq 0}$  of positive numbers such that  $\lim_{n \rightarrow +\infty} \gamma_n = 0$ :*

$$(4.7) \quad \mathbb{P}_\mu^n \left\{ \left| \frac{N_{n,\gamma_n}}{2nF_\nu(\gamma_n)} - 1 \right| > \varepsilon \right\} \leq 2 \exp(-\varepsilon^2 n F_\nu(\gamma_n)/6),$$

$$(4.8) \quad \mathbb{P}_\mu^n \left\{ \left| \frac{\overline{K}_{n,\rho,\gamma_n}}{2nF_\nu(\gamma_n)} - c_K \right| > \varepsilon \right\} \leq 2 \exp(-C_1 \varepsilon^2 n F_\nu(\gamma_n)),$$

where  $c_K \triangleq \int K(y) \exp(y) dy$  and  $C_1 \triangleq C_1(K) = (3 \int K^2(y) \exp(y) dy + 1/3)^{-1}/8$ .

Then in view of proposition 4.3 and proposition 4.4 we define the optimal bandwidth  $h = h_n$  as in the definition 3.4:

**Definition 4.5.** Let  $h_n$  be defined as the smallest solution (well defined in view of proposition 4.6) of the equation:

$$(4.9) \quad \omega(h) = \frac{\sigma}{\sqrt{nF_\nu(h)}},$$

where  $F_\nu(h) = \int_0^h \nu(t) dt$ .

The following proposition gives the existence and the uniqueness of  $h_n$ . Moreover it tells that  $h_n$  is "large", since it converges slowly to 0.

**Proposition 4.6.** *Under assumption G there exist a solution to equation 4.9 for  $n$  large enough. We have  $\lim_{n \rightarrow +\infty} h_n = 0$  and:*

$$(4.10) \quad h_n = \varphi(n^{-1}),$$

where  $\varphi$  belongs to the de Haan class  $\text{IIV}_0(\ell)$  (see definition B.5) with  $\ell$  a slowly varying function depending on  $\nu$ . Moreover, since  $\text{IIV}_0(\ell) \subset \text{RV}_0(0)$  (see lemma B.6),  $\varphi$  is in particular slowly varying.

Let us define the rate

$$(4.11) \quad r_n \triangleq \omega(h_n),$$

where  $h_n$  is given by definition 4.5. Then we define  $\widehat{f}_n(x_0) \triangleq \widehat{f}_n(x_0, h_n)$  and  $\widetilde{f}_n(x_0) \triangleq \widetilde{f}_n(x_0, h_n)$ , with  $\eta_n = r_n^\tau$  where  $\tau > 0$ , and  $\alpha_n = r_n^{-\gamma}$  where  $\gamma > 0$ .

**Theorem 4.7.** *Under assumptions G, K one has for any  $p > 0$ :*

$$(4.12) \quad \sup_{f \in \Sigma_{h_n, \alpha_n}(x_0, \omega)} \mathbb{E}_{f, \mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \} \leq C(p) + o_n(1),$$

$$(4.13) \quad \sup_{f \in \Sigma_{h_n, \alpha_n}(x_0, \omega)} \mathbb{E}_{f, \mu}^n \{ r_n^{-p} |\widetilde{f}_n(x_0) - f(x_0)|^p \} \leq C(K, p) + o_n(1),$$

where  $r_n$  is given by (4.11),  $C(K, p)$  is a constant depending on  $K$  and  $p$  only,  $C(p)$  is the same as in theorem 3.6. Moreover we have:

$$r_n = \ell_{\omega, \nu}(n^{-1}),$$

where  $\ell_{\omega, \nu}$  is a slowly varying function such that  $\lim_{h \rightarrow 0^+} \ell_{\omega, \nu}(h) = 0$ .

### 4.3. The lower bound.

**Theorem 4.8.** *Under assumption G, one has for any  $p > 0$  and for  $n$  large enough:*

$$(4.14) \quad \mathcal{R}_{n,p}(\Sigma_{h_n, \alpha_n}(x_0, \omega), \mu, x_0) \geq C_5 r_n,$$

where  $r_n$  is defined in (4.11) and  $C_5$  is some constant depending on  $s$  and  $p$  only.

Then  $r_n$  is the minimax rate of convergence over the class  $\Sigma_{h_n, \alpha_n}(x_0, \omega)$ .

**4.4. Example.** We consider a density  $\mu$  satisfying G with  $\nu(h) = \exp(-1/h^\alpha)$ , for  $\alpha > 0$  and we take  $\omega(h) = rh^s$  for  $r > 0$  and  $0 < s \leq 1$ . It is an easy computation to see that  $\nu$  belongs to the class  $\Gamma V_0(\rho)$  for the auxiliary function  $\rho(h) = \alpha^{-1}h^{\alpha+1}$ . In this case, we find the following minimax rate (see section 5.2.1 for the details):

$$r_n \asymp r(\log n)^{-s/\alpha}.$$

As told by theorem 4.7 and 4.8, we find that the minimax rate in this example is very slow. We note that the parameters  $s$  and  $\alpha$  are on the same scale.

## 5. PROOFS

We begin with the proofs of the results on the local polynomial estimation procedure in the considered model, without assumptions on the density of the design. First we give the proof of proposition 2.3. In all the following,  $\langle \cdot, \cdot \rangle$  denotes the Euclidean scalar product on  $\mathbb{R}^{k+1}$ ,  $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{k+1}$ ,  $\|\cdot\|_\infty$  stands for the sup norm in  $\mathbb{R}^{k+1}$  and  $\|\cdot\|$  stands for the Euclidean norm in  $\mathbb{R}^{k+1}$ .

*Proof of proposition 2.3.* On  $\Omega_h^K$  we have in view of definition 2.1 that  $\widetilde{\mathbf{X}}_h^K = \mathbf{X}_h^K$  and  $\mathbf{X}_h^K$  is invertible. Let  $0 < \varepsilon \leq 1/2$ , and  $n \geq 1$ . We can find a polynomial  $P_f^{n, \varepsilon}$  of order  $k$  such that

$$\sup_{|x-x_0| \leq h} |f(x) - P_f^{n, \varepsilon}(x)| \leq \omega_{f,k}(x_0, h) + \frac{\varepsilon}{\sqrt{n}}.$$

In particular with  $h = 0$  we get  $|f(x_0) - P_f^{n, \varepsilon}(x_0)| \leq \frac{\varepsilon}{\sqrt{n}}$ . Defining  $\theta_h \in \mathbb{R}^{k+1}$  such that  $P_f^{n, \varepsilon} = P_{\theta_h}$  (see (2.6)) we get

$$|\widehat{f}_h(x_0) - f(x_0)| \leq \frac{\varepsilon}{\sqrt{n}} + |\langle \widehat{\theta}_h - \theta_h, e_1 \rangle| = \frac{\varepsilon}{\sqrt{n}} + |\langle (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^K (\widehat{\theta}_h - \theta_h), e_1 \rangle|.$$

Then we have for  $j \in \{0, \dots, k\}$  by (2.4) and (1.1):

$$\begin{aligned} (\mathbf{X}_h^K(\widehat{\theta}_h - \theta_h))_j &= \langle \widehat{f}_h - P_f^{n,\varepsilon}, \phi_{j,h} \rangle_{h,K} = \langle Y - P_f^{n,\varepsilon}, \phi_{j,h} \rangle_{h,K} \\ &= \langle f - P_f^{n,\varepsilon}, \phi_{j,h} \rangle_{h,K} + \langle Y - f, \phi_{j,h} \rangle_{h,K} \\ &= \langle f - P_f^{n,\varepsilon}, \phi_{j,h} \rangle_{h,K} + \langle \xi, \phi_{j,h} \rangle_{h,K} \\ &\triangleq B_{h,j} + V_{h,j}, \end{aligned}$$

thus  $\mathbf{X}_h^K(\widehat{\theta}_h - \theta_h) = B_h + V_h$ . In view of assumption K, we get:

$$|B_{h,j}| = |\langle f - P_f^{n,\varepsilon}, \phi_{j,h} \rangle_{h,K}| \leq \|f - P_f^{n,\varepsilon}\|_{h,K} \|\phi_{j,h}\|_{h,K} \leq N_{n,h} K_\infty (\omega_{f,k}(x_0, h) + \frac{\varepsilon}{\sqrt{n}}),$$

thus  $\|B_h\|_\infty \leq N_{n,h} K_\infty (\omega_{f,k}(x_0, h) + \frac{\varepsilon}{\sqrt{n}})$ . Moreover, since  $\lambda^{-1}(\mathcal{X}_h) \leq N_{n,h}^{1/2} \leq n^{1/2}$  on  $\Omega_{h,K}$ , we have:

$$\begin{aligned} | \langle (\mathbf{X}_h^K)^{-1} B_h, e_1 \rangle | &\leq \|(\mathbf{X}_h^K)^{-1}\| \|B_h\| \leq \|(\mathbf{X}_h^K)^{-1}\| \sqrt{k+1} \|B_h\|_\infty \\ &\leq \lambda^{-1}(\mathcal{X}_h^K) \sqrt{k+1} K_\infty \omega_{f,k}(x_0, h) + \sqrt{k+1} K_\infty \varepsilon, \end{aligned}$$

where we last used the fact that  $\|M^{-1}\| = \lambda^{-1}(M)$  for a positive symmetrical matrix. The variance term  $V_h$  is clearly conditional on  $\mathfrak{X}_n$  a centered Gaussian vector, and its covariance matrix is equal to  $\sigma^2 \mathbf{X}_h^{K^2}$ . Thus the random variable  $\langle (\mathbf{X}_h^K)^{-1} V_h, e_1 \rangle_{h,K}$  is, conditional on  $\mathfrak{X}_n$ , centered Gaussian, of variance:

$$\begin{aligned} v_h^2 &= \sigma^2 \langle e_1, (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^{K^2} (\mathbf{X}_h^K)^{-1} e_1 \rangle \leq \sigma^2 \langle e_1, (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^K (\mathbf{X}_h^K)^{-1} e_1 \rangle \\ &= \sigma^2 \langle e_1, (\mathbf{X}_h^K)^{-1} e_1 \rangle \\ &\leq \sigma^2 \|(\mathbf{X}_h^K)^{-1}\|^2 = \sigma^2 N_{n,h}^{-1} \lambda^{-1}(\mathcal{X}_h^K), \end{aligned}$$

since  $K \leq 1$ . Then

$$\lambda(\mathcal{X}_h^K) = \inf_{\|x\|=1} \langle x, \mathcal{X}_h^K x \rangle \leq \|\mathcal{X}_h^K e_1\| \leq \sqrt{k+1},$$

since  $\mathcal{X}_h^K$  is symmetrical and its entries are smaller than 1 in absolute value. Thus:

$$v_h^2 \leq \sigma^2 N_{n,h}^{-1} \lambda^{-1}(\mathcal{X}_h^K) \leq \sigma^2 N_{n,h}^{-1} (k+1) \lambda^{-2}(\mathcal{X}_h^K),$$

and the proposition follows.  $\square$

*Proof of proposition 2.4.* (2)  $\Rightarrow$  (1): In view of assumption M one has for  $n$  large enough  $\mathbb{E}_\mu^n \{N_{n,C\gamma_n}\} = 2n \int_0^{C\gamma_n} \nu(x) dx = 2n F_\nu(C\gamma_n)$  thus (2) entails  $2n \lambda_n^{-1} F_\nu(C\gamma_n) \sim \phi(C)$  as  $n \rightarrow +\infty$  and then  $F_\nu \in \text{RV}_0$  in view of theorem A.10. Since  $F_\nu(0) = 0$  we have more precisely  $F_\nu \in \text{RV}_0(\alpha)$  for  $\alpha \geq 0$  and since  $\nu$  is monotone, using proposition A.6 we have  $\nu \in \text{RV}_0(\alpha - 1)$ .

(3)  $\Rightarrow$  (2): Let  $\varepsilon > 0$ . We define the event

$$A_n(C, \varepsilon) = \left\{ \left| \frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1 \right| \leq \varepsilon \right\}.$$

Then:

$$\lambda_n^{-1} \mathbb{E}_\mu^n \{N_{n,C\gamma_n}\} = \lambda_n^{-1} \mathbb{E}_\mu^n \{N_{n,C\gamma_n} (\mathbf{1}_{A_n(C,\varepsilon)} + \mathbf{1}_{A_n^c(C,\varepsilon)})\} \leq (1 + \varepsilon) \phi(C) + n \lambda_n^{-1} \mathbb{P}_\mu^n \{A_n^c(C, \varepsilon)\},$$

and then  $\limsup_n \lambda_n^{-1} \mathbb{E}_\mu^n \{N_{n,C\gamma_n}\} \leq (1 + \varepsilon) \phi(C)$ . On the other side:

$$\lambda_n^{-1} \mathbb{E}_\mu^n \{N_{n,C\gamma_n}\} \geq \lambda_n^{-1} \mathbb{E}_\mu^n \{N_{n,C\gamma_n} \mathbf{1}_{A_n(C,\varepsilon)}\} \geq (1 - \varepsilon) \phi(C) \mathbb{P}_\mu^n \{A_n(C, \varepsilon)\},$$

and then  $\liminf_n \lambda_n^{-1} \mathbb{E}_\mu^n \{N_{C\gamma_n}\} \geq (1 - \varepsilon)\phi(C)$ .

(1)  $\Rightarrow$  (3): Let  $\nu \in \text{RV}_0(\beta)$  and  $0 < \varepsilon \leq 1/2$ . If  $\beta > -1$  we define  $\gamma_n = n^{-1/(2(\beta+1))}$ , and theorem A.4 entails  $F_\nu \in \text{RV}_0(\beta + 1)$ , thus we can write  $F_\nu(h) = h^{\beta+1}\ell_\nu(h)$  where  $\ell_\nu$  is slowly varying. If  $\beta = -1$  we define  $\gamma_n = n^{-1}$ , and proposition A.5 entails again  $F_\nu \in \text{RV}_0(\beta + 1)$ . We note that  $\lim_n \gamma_n = 0$  and  $\gamma_{n+1} \sim \gamma_n$  as  $n \rightarrow +\infty$ . In view of proposition 3.3 we get for  $n$  large enough:

$$\mathbb{P}_\mu^n \left\{ \left| \frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1 \right| > \varepsilon \right\} \leq 2 \exp(-C_1 \varepsilon^2 \phi(C)\lambda_n/4),$$

where we used  $\ell_\nu$  is slowly varying and where we defined  $\lambda_n \triangleq 2nF_\nu(\gamma_n)$  and  $\phi(C) \triangleq C^{\beta+1}$ . Then we clearly have  $\lim_n n\lambda_n^{-1} = +\infty$  (when  $\beta = -1$ ,  $\lim_n n\lambda_n^{-1} = \lim_n \ell_\nu^{-1}(n^{-1}) = +\infty$ ), and the proposition follows.  $\square$

**5.1. Regularly varying design.** We define  $K_0(x) = \frac{1}{2}\mathbf{1}_{|x| \leq 1}$  the rectangular kernel.

*Proof of proposition 3.3.* This proposition is actually a corollary of lemma 5.1 below, for the choice  $K = K_0$  and  $\alpha = 0$ . When  $\beta > -1$ , the Karamata theorem A.4 entails  $F_\nu(\gamma_n) \sim (1 + \beta)^{-1}\gamma_n\nu(\gamma_n)$  as  $n \rightarrow +\infty$ , and entail together with (5.2) the result. When  $\beta = -1$ , the result is straightforward in view of (5.3) for the choice  $K = K_0$ .  $\square$

*Proof of proposition 3.5.* Let us define  $G(h) = \omega^2(h)F_\nu(h)$ . When  $\beta > -1$ , the Karamata theorem A.4 entails  $F_\nu \in \text{RV}_0(\beta + 1)$ . When  $\beta = -1$ , the proposition A.5 entails  $F_\nu$  is slowly varying. Thus, in view of proposition A.2,  $G$  belongs to  $\text{RV}_0(1 + 2s + \beta)$  for any  $\beta \geq -1$ . The function  $G$  is continuous and such that  $\lim_{h \rightarrow 0^+} G(h) = 0$  in view of proposition A.3, since  $1 + 2s + \beta > 0$ . Then for  $n$  large enough  $h_n$  is well defined and given by  $h_n = G^\leftarrow(\sigma^2/n)$ , where  $G^\leftarrow(h) \triangleq \inf\{y \geq 0 \mid G(y) \geq h\}$  is the generalised inverse of  $G$ . Theorem A.9 entails  $G^\leftarrow \in \text{RV}_0(1/(1 + 2s + \beta))$ , thus there is a slowly varying function  $\ell_{\omega,\nu}$  such that  $G^\leftarrow(h) = h^{1/(1+2s+\beta)}\ell_{\omega,\nu}(h)$ . Thus:

$$h_n = \sigma^{2/(1+2s+\beta)}n^{-1/(1+2s+\beta)}\ell_{\omega,\nu}(\sigma^2/n) \sim \sigma^{2/(1+2s+\beta)}n^{-1/(1+2s+\beta)}\ell_{\omega,\nu}(1/n) \text{ as } n \rightarrow +\infty,$$

where we used  $\ell_{\omega,\nu}$  is slowly varying. When  $\omega(h) = rh^s$ , we can write more precisely  $h_n = G^\leftarrow(\sigma^2/(r^2n))$ , where  $G(h) = h^{2s}F_\nu(h)$ , so (3.3) and (3.4) follow. The asymptotic equivalence of any solutions  $h_{n,1}$  and  $h_{n,2}$  is given by the theorem A.9.  $\square$

**Study of the term  $\lambda(\mathcal{X}_{h_n}^K)$ .** We recall that the matrix  $\mathcal{X}_h^K$  is defined as the symmetrical and non-negative matrix with entries, for  $0 \leq j, l \leq k$ ,  $(\mathcal{X}_h^K)_{j,l} = \overline{K}_{n,h,j+l}$ , where:

$$\overline{K}_{n,h,\alpha} \triangleq N_{n,h}^{-1} \sum_{i=1}^n \left( \frac{X_i - x_0}{h} \right)^\alpha K \left( \frac{X_i - x_0}{h} \right),$$

for  $\alpha \in \mathbb{N}$ . Let us define  $K_{n,h,\alpha} \triangleq N_{n,h}\overline{K}_{n,h,\alpha}$ , and:

$$(5.1) \quad K_{\alpha,\beta} \triangleq (1 + (-1)^\alpha) \int_0^1 y^{\alpha+\beta} K(y) dy.$$

We define, for any  $\varepsilon > 0$ , the event

$$D_{n,\alpha,K,\varepsilon} \triangleq \left\{ \left| \frac{K_{n,\gamma_n,\alpha}}{n\gamma_n\nu(\gamma_n)} - K_{\alpha,\beta} \right| \leq \varepsilon \right\}.$$

**Lemma 5.1.** For any  $\alpha \in \mathbb{N}$  and  $0 < \varepsilon \leq 1$ , one has under assumptions  $K, R$  with  $\beta > -1$  that for any sequence  $(\gamma_n)_n$  of positive numbers converging to 0:

$$(5.2) \quad \mathbb{P}_\mu^n \{D_{n,\alpha,K,\varepsilon}^c\} \leq 2 \exp(-C_1 \varepsilon^2 n \gamma_n \nu(\gamma_n)),$$

for  $n$  large enough, where  $C_1 = C_1(\alpha, \beta, K) = (3 \int_0^1 y^{2\alpha+\beta} K^2(y) dy + K_\infty/3)^{-1}/8$ . When  $\beta = -1$  we have:

$$(5.3) \quad \mathbb{P}_\mu^n \left\{ \left| \frac{K_{n,\gamma_n,0}}{n F_\nu(\gamma_n)} - 2K(0) \right| > \varepsilon \right\} \leq 2 \exp(-C_2 \varepsilon^2 n F_\nu(\gamma_n)),$$

where  $C_2 = C_2(K) = (2 + K_\infty/3)^{-1}/8$ .

*Proof.* Let  $\alpha \in \mathbb{N}$ ,  $0 < \varepsilon \leq 1$ . We first prove the result when  $\beta > -1$ . We define  $Q_{i,n,\alpha} \triangleq \left(\frac{X_i - x_0}{\gamma_n}\right)^\alpha K\left(\frac{X_i - x_0}{\gamma_n}\right)$ ,  $Z_{i,n,\alpha} \triangleq Q_{i,n,\alpha} - \mathbb{E}_\mu^n \{Q_{i,n,\alpha}\}$ . Under assumption R one has for  $i \in \{1, \dots, n\}$ :

$$\frac{1}{\gamma_n \nu(\gamma_n)} \mathbb{E}_\mu^n \{Q_{i,n,\alpha}\} = \frac{1 + (-1)^\alpha}{\ell_\nu(\gamma_n)} \int_0^1 y^{\alpha+\beta} K(y) \ell_\nu(y \gamma_n) dy,$$

where we used assumption K and the fact that for  $n$  large enough  $[x_0 - \gamma_n, x_0 + \gamma_n] \subset W$ . Now theorem A.7 entails:

$$\lim_{n \rightarrow +\infty} \frac{1}{\ell_\nu(\gamma_n)} \int_0^1 y^{\alpha+\beta} K(y) \ell_\nu(y \gamma_n) dy = \int_0^1 y^{\alpha+\beta} K(y) dy.$$

Then for  $n$  large enough:

$$(5.4) \quad D_{n,\alpha,K,\varepsilon}^c \subset \left\{ \left| \frac{1}{n \gamma_n \nu(\gamma_n)} \sum_{i=1}^n Z_{i,n,\alpha} \right| > \varepsilon/2 \right\}.$$

We have  $\mathbb{E}_\mu^n \{Z_{i,n,\alpha}\} = 0$ ,  $|Z_{i,n,\alpha}| \leq 2K_\infty$  in view of assumption K. We get in the same way as previously

$$b_{n,\alpha}^2 \triangleq \sum_{i=1}^n \mathbb{E}_\mu^n \{Z_{i,n,\alpha}^2\} \leq n \mathbb{E}_\mu^n \{Q_{i,n,\alpha}^2\} \sim 2n \gamma_n \nu(\gamma_n) \int_0^1 y^{2\alpha+\beta} K^2(y) dy \text{ as } n \rightarrow +\infty,$$

then for  $n$  large enough we have  $b_{n,\alpha}^2 \leq C_2 n \gamma_n \nu(\gamma_n)$ , where  $C_2 \triangleq 3 \int_0^1 y^{2\alpha+\beta} K^2(y) dy > 0$ . Since the  $Z_{i,n,\alpha}$  are independants we can apply the Bernstein inequality. If  $\tau_n \triangleq \frac{\varepsilon}{2} n \gamma_n \nu(\gamma_n)$ , equation (5.4) and the Bernstein inequality entails:

$$\mathbb{P}_\mu^n \{D_{n,\alpha,K,\varepsilon}^c\} \leq 2 \exp\left(\frac{-\tau_n^2}{2(b_{n,\alpha}^2 + 2K_\infty \tau_n/3)}\right) \leq 2 \exp(-C_1 \varepsilon^2 n \gamma_n \nu(\gamma_n)),$$

thus (5.2) follows. When  $\beta = -1$ ,  $\nu(t) = t^{-1} \ell_\nu(t)$ , and we procede as previously. We define  $Z_{i,n} \triangleq Q_{i,n,0} - \mathbb{E}_{f,\mu}^n \{Q_{i,n,0}\}$ . We have in view of proposition A.8:

$$\lim_{n \rightarrow +\infty} \frac{1}{F_\nu(\gamma_n)} \mathbb{E}_\mu^n \{Q_{i,n,0}\} = \lim_{n \rightarrow +\infty} \frac{2}{F_\nu(\gamma_n)} \int_0^1 K(t/h) \ell_\nu(t) dt/t = 2K(0) > 0.$$

Then for  $n$  large enough one has

$$\left\{ \left| \frac{K_{n,\gamma_n,0}}{n F_\nu(\gamma_n)} - 2K(0) \right| > \varepsilon \right\} \subset \left\{ \left| \frac{1}{n F_\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n} \right| > \varepsilon/2 \right\}.$$



The  $Z_{i,n}$  are independent and centered variables, and  $|Z_{i,n}| \leq 2K_\infty$ . Moreover, in view of assumption K, we have

$$b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}_\mu^n \{Z_{i,n}^2\} \leq n \mathbb{E}_\mu^n \{Q_{i,n,0}^2\} = 2n \int_0^1 K^2(t/\gamma_n) \nu(t) dt \leq 2n F_\nu(\gamma_n).$$

Then using again the Bernstein inequality we get (5.3).  $\square$

The following proposition is a deviation result for the difference between the smallest eigenvalue of the matrix  $\mathcal{X}_h^K$  and the smallest eigenvalue  $\lambda_\beta^K > 0$  of the positive matrix  $\mathcal{X}_\beta^K$  (defined in (3.7)). This proposition entails that the matrix  $\mathcal{X}_{h_n}^K$  is non-degenerate with a large probability when  $n$  is large.

**Proposition 5.2.** *Under assumption R with  $\beta > -1$ , if  $\gamma_n$  is a sequence of positive numbers converging to 0, and  $0 < \varepsilon \leq 1$ , then one has for  $n$  large enough:*

$$\mathbb{P}_\mu^n \{|\lambda(\mathcal{X}_{\gamma_n}^K) - \lambda_\beta^K| > \varepsilon\} \leq 4(k+1) \exp(-C(\beta)\eta^2 n \gamma_n \nu(\gamma_n)),$$

with  $\lambda_\beta^K = \lambda(\mathcal{X}_\beta^K) > 0$ ,  $C(\beta) = (3/(\beta+1) + 1/3)^{-1}/8$  and  $\eta \triangleq \frac{\varepsilon}{2(\beta+1)(k+1)^2 + \varepsilon}$ .

*Proof.* Since  $\lambda_\beta^K$  is the smallest eigenvalue of  $\mathcal{X}_\beta^K$  we have  $\lambda_\beta^K > 0$ , otherwise, defining  $\mathbf{p}(y) = (1, y, \dots, y^k)$  and since  $\mathcal{X}_\beta^K$  is symmetrical, we have

$$0 = \lambda_\beta^K = \inf_{\|x\|=1} \langle x, \mathcal{X}_\beta^K x \rangle = \langle x_0, \mathcal{X}_\beta^K x_0 \rangle = \int_{-1}^1 ({}^t x_0 \mathbf{p}(y))^2 y^\beta K(y) dy,$$

where  $x_0 \neq 0$  is the normalized eigenvector associated to the eigenvalue  $\lambda_\beta^K$ , and where we used the fact that  $\lambda(M) = \inf_{\|x\|=1} \langle x, Mx \rangle$  for any symmetrical matrix  $M$ . Then  $\forall y \in \text{Supp } K$  we have  ${}^t x_0 \mathbf{p}(y) = 0$ , which leads to a contradiction since  $y \mapsto {}^t x_0 \mathbf{p}(y)$  is a polynomial. For any  $\varepsilon > 0$  we introduce the events:

$$(5.5) \quad \mathbf{A}_{n,\varepsilon} = \{|\lambda(\mathcal{X}_{\gamma_n}^K) - \lambda_\beta^K| \leq \varepsilon\}, \quad \mathbf{B}_{n,\alpha,\varepsilon} = \left\{ \left| \overline{K}_{n,\gamma_n,\alpha} - \frac{\beta+1}{2} K_{\alpha,\beta} \right| \leq \varepsilon \right\}.$$

We note that

$$(5.6) \quad \bigcap_{\alpha=0}^{2k} \mathbf{B}_{n,\alpha,\varepsilon/(k+1)^2} \subset \mathbf{A}_{n,\varepsilon},$$

for  $0 < \varepsilon \leq 1$ . Noting that  $\mathbf{D}_{n,0,K_0,\varepsilon_1} = \left\{ \left| \frac{N_{n,\gamma_n}}{2n\gamma_n\nu(\gamma_n)} - \frac{1}{\beta+1} \right| \leq \varepsilon_1 \right\}$ , we have

$$\mathbf{D}_{n,0,K_0,\varepsilon_1} \cap \mathbf{D}_{n,\alpha,K,\varepsilon_1} \subset \mathbf{B}_{n,\alpha,\varepsilon},$$

and for any  $\alpha \in \mathbb{N}$ , where  $\varepsilon_1 \triangleq \frac{\varepsilon}{2(\beta+1)+\varepsilon}$ . Now using (5.6) we get for  $\eta \triangleq \frac{\varepsilon}{2(\beta+1)(k+1)^2 + \varepsilon}$ :

$$\mathbf{D}_{n,0,K_0,\eta} \cap \bigcap_{\alpha=0}^{2k} \mathbf{D}_{n,\alpha,K,\eta} \subset \mathbf{A}_{n,\varepsilon}.$$

Finally, lemma 5.1 entails for  $n$  large enough

$$\mathbb{P}_\mu^n \{\mathbf{A}_{n,\varepsilon}^c\} \leq \mathbb{P}_\mu^n \{\mathbf{D}_{n,0,K_0,\eta}^c\} + \sum_{\alpha=0}^{2k} \mathbb{P}_\mu^n \{\mathbf{D}_{n,\alpha,K,\eta}^c\} \leq 4(k+1) \exp(-C(\beta)\eta^2 n \gamma_n \nu(\gamma_n)). \quad \square$$

**Proof of the upper bound.**

**Proposition 5.3.** *For any  $p > 0$  and  $h > 0$  such that  $N_{n,h} > 0$ , the estimator  $\widehat{f}_h$  (see definition 2.1) satisfies:*

$$\sup_{f \in \mathcal{U}(x_0, h, \alpha_n)} \mathbb{E}_{f, \mu}^n \{ |\widehat{f}_h(x_0)|^p \} = O((\sqrt{n}\alpha_n)^p).$$

*Proof.* Using the fact that  $\lambda(A + B) \geq \lambda(A) + \lambda(B)$  when  $A$  and  $B$  are symmetrical and non-negative matrices, we get  $\lambda(\widetilde{\mathbf{X}}_h^K) \geq N_{n,h}^{1/2} > 0$ , and then  $\widetilde{\mathbf{X}}_h^K$  is invertible. Thus equation (2.9) entails  $|\widehat{f}_h(x_0)| = | \langle (\widetilde{\mathbf{X}}_h^K)^{-1} \widetilde{\mathbf{X}}_h^K \widehat{\theta}_h, e_1 \rangle | = | \langle (\widetilde{\mathbf{X}}_h^K)^{-1} \mathbf{Y}_h, e_1 \rangle |$ . In view of (1.1) we can decompose for  $j \in \{0, \dots, k\}$ :

$$(\mathbf{Y}_h)_j = \langle Y, \phi_{j,h} \rangle_{h,K} = \langle f, \phi_{j,h} \rangle_{h,K} + \langle \xi, \phi_{j,h} \rangle_{h,K} \triangleq B_{h,j} + V_{h,j}.$$

Since  $f \in \mathcal{U}(x_0, h, \alpha_n)$  we have under assumption K that  $|B_{h,j}| \leq \alpha_n N_{n,h} K_\infty$ , thus  $\|B_h\|_\infty \leq \alpha_n N_{n,h} K_\infty$ . As in the proof of proposition 2.3 we have that  $\langle (\widetilde{\mathbf{X}}_h^K)^{-1} V_h, e_1 \rangle$  is, conditional on  $\mathfrak{X}_n$ , a centered Gaussian variable, of variance:

$$\begin{aligned} v_h^2 &= \sigma^2 \langle e_1, (\widetilde{\mathbf{X}}_h^K)^{-1} \mathbf{X}_h^{K^2} (\widetilde{\mathbf{X}}_h^K)^{-1} e_1 \rangle \leq \sigma^2 \langle e_1, (\widetilde{\mathbf{X}}_h^K)^{-1} \mathbf{X}_h^K (\widetilde{\mathbf{X}}_h^K)^{-1} e_1 \rangle \\ &\leq \sigma^2 \|(\widetilde{\mathbf{X}}_h^K)^{-1}\|^2 \|\mathbf{X}_h^K\|. \end{aligned}$$

Assumption K entails all the elements of the matrix  $\mathbf{X}_h^K$  are smaller than  $N_{n,h} K_\infty$ , thus  $\|\mathbf{X}_h^K\| \leq (k+1)N_{n,h} K_\infty$ . Since  $\widetilde{\mathbf{X}}_h^K$  is symmetrical we get  $\|(\widetilde{\mathbf{X}}_h^K)^{-1}\| = \lambda^{-1}(\widetilde{\mathbf{X}}_h^K) \leq N_{n,h}^{-1/2}$ , and then  $v_h^2 \leq K_\infty \sigma^2 (k+1)$ . Finally, we have

$$\begin{aligned} |\widehat{f}_h(x_0)| &\leq | \langle (\widetilde{\mathbf{X}}_h^K)^{-1} B_h, e_1 \rangle | + | \langle (\widetilde{\mathbf{X}}_h^K)^{-1} V_h, e_1 \rangle | \leq \|(\widetilde{\mathbf{X}}_h^K)^{-1}\| \|B_h\| + \sigma \sqrt{k+1} |\gamma_h| \\ &\leq \sqrt{k+1} (\alpha_n \sqrt{n} + \sigma |\gamma_h|), \end{aligned}$$

where  $\gamma_h$  is, conditional on  $\mathfrak{X}_n$ , centered Gaussian with variance smaller than 1. The result follows by integration.  $\square$

*Proof of theorem 3.6.* Let  $0 < \varepsilon \leq 1$ . Let assume  $\beta > -1$ . Let define the event

$$\mathcal{A}_{n,\varepsilon} \triangleq \{ |\lambda(\mathcal{X}_{h_n}^K) - \lambda_\beta^K| \leq \varepsilon \} \cap \left\{ \left| \frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1 \right| \leq \varepsilon \right\}.$$

Proposition 5.2 and proposition 3.3 together entails:

$$\begin{aligned} \mathbb{P}_\mu^n \{ \mathcal{A}_{n,\varepsilon}^c \} &\leq 2 \exp(-C(\beta)\varepsilon^2 n F_\nu(h_n)) + 2(k+2) \exp\{-C(\beta)\eta^2 n h_n \nu(h_n)\} \\ &\leq 2(k+3) \exp(-C(\beta)\eta^2 r_n^{-2}/2), \end{aligned}$$

for  $n$  large enough, where we used the Karamata theorem and  $r_n^2 = \omega^2(h_n) = \sigma^2/(nF_\nu(h_n))$ . Since  $\lim_n nF_\nu(h_n) = +\infty$ , we have for  $n$  large enough  $\mathcal{A}_{n,\varepsilon} \subset \{ \lambda(\mathbf{X}_{h_n}^K) > N_{n,h_n}^{1/2} \}$  and in particular on the event  $\mathcal{A}_{n,\varepsilon}$  the matrix  $\mathbf{X}_{h_n}^K$  is invertible. Then using proposition 2.3 and since  $f \in \mathcal{F}_{h_n}(x_0, \omega)$ , we get:

$$\begin{aligned} |\widehat{f}_n(x_0) - f(x_0)| \mathbf{1}_{\mathcal{A}_{n,\varepsilon}} &\leq (\lambda_\beta^K - \varepsilon)^{-1} \sqrt{k+1} K_\infty (\omega(h_n) + \frac{\sigma}{\sqrt{(2-\varepsilon)nF_\nu(h_n)}} |\gamma_{h_n}|) \\ &\leq (\lambda_\beta^K - \varepsilon)^{-1} \sqrt{k+1} K_\infty \omega(h_n) (1 + |\gamma_{h_n}|), \end{aligned}$$

where we last used the definition 3.4 of  $h_n$ . Since  $\gamma_{h_n}$  is conditional on  $\mathfrak{X}_n$  centered Gaussian such that  $\mathbb{E}_{f,\mu}^n \{ \gamma_{h_n}^2 | \mathfrak{X}_n \} \leq 1$ , we get for any  $p > 0$  and  $n$  large enough:

$$\sup_{f \in \mathcal{F}_{h_n}(x_0, \omega)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}} | \mathfrak{X}_n \} \leq (\lambda_\beta^K - \varepsilon)^{-p} (k+1)^{p/2} K_\infty^p C(p),$$

where  $C(p)$  is defined in section 3.2. Now we work on the complementary event  $\mathcal{A}_{n,\varepsilon}^c$ . We recall that  $\alpha_n = n^\gamma$  for some  $\gamma > 0$ . If  $N_{n,h_n} = 0$  we have by definition  $\widehat{f}_n(x_0) = 0$  and then

$$\sup_{f \in \mathcal{U}(x_0, h_n, \alpha_n)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}^c} \} \leq r_n^{-p} n^{p\gamma} \mathbb{P}_{f,\mu}^n \{ \mathcal{A}_{n,\varepsilon}^c \} = o_n(1).$$

Then we assume  $N_{n,h_n} > 0$ . the proposition 5.3 entails:

$$\begin{aligned} \sup_{f \in \mathcal{U}(x_0, h_n, \alpha_n)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \} &\leq 2^p r_n^{-p} (\sqrt{\mathbb{E}_{f,\mu}^n \{ |\widehat{f}_n(x_0)|^{2p} \}} + n^{p\gamma}) \sqrt{\mathbb{P}_{f,\mu}^n \{ \mathcal{A}_{n,\varepsilon}^c \}} \\ &\leq O(r_n^{-p} n^{p(1+2\gamma)}) \sqrt{\mathbb{P}_{f,\mu}^n \{ \mathcal{A}_{n,\varepsilon}^c \}} = o_n(1). \end{aligned}$$

Then we have proved inequality (3.8). Since  $\omega \in \text{RV}_0(s)$  and  $G^\leftarrow \in \text{RV}_0(s/(1+2s+\beta))$  (see the proof of proposition 3.5) we have in view of proposition A.2:  $\omega \circ G^\leftarrow \in \text{RV}_0(s/(1+2s+\beta))$ . Thus we get  $r_n = \omega(h_n) = \omega(G^\leftarrow(\sigma^2/n))$  and:

$$r_n = \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(\sigma^2/n) \sim \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(1/n) \text{ as } n \rightarrow +\infty.$$

When assumption R holds with  $\beta = -1$  and  $0 < s \leq 1$ , the bias variance equation (2.12) becomes:

$$|\widehat{f}_n(x_0) - f(x_0)| \leq (\overline{K}_{n,h_n,0})^{-1} K_\infty(\omega(h_n) + \sigma N_{n,h_n}^{-1/2} |\gamma h_n|).$$

We define the event  $\mathcal{A}_{n,\varepsilon} = \{ |\frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1| \leq \varepsilon \} \cap \{ |\frac{K_{n,h_n,0}}{2nF_\nu(h_n)} - K(0)| \leq \varepsilon \}$ , so (5.3) in lemma 5.1 and proposition 3.3 together entail  $\mathbb{P}_{f,\mu}^n \{ \mathcal{A}_{n,\varepsilon}^c \} \leq 4 \exp(-\varepsilon^2 C(\beta) n F_\nu(h_n))$ . The rest of the proof is straightforward.  $\square$

**Proof of the lower bound.** Since we are considering a pointwise risk, we know (see for instance Tsybakov (2003)) that a 2-hypothesis reduction scheme is enough to prove the lower bound. The following proposition can be found in Tsybakov (2003).

**Proposition 5.4.** *If there are 2 elements  $f_0$  and  $f_1$  of a class  $\Sigma$ , such that the Kullback-Leibler distance between the corresponding probabilities  $\mathbb{P}_0$  and  $\mathbb{P}_1$  satisfies  $\mathcal{K}(\mathbb{P}_0, \mathbb{P}_1) < Q < +\infty$  with  $|f_0(x_0) - f_1(x_0)| \geq 2cr_n$  for some constant  $c > 0$ , then the pointwise minimax risk  $\mathcal{R}_{n,p}(\Sigma, \mu, x_0)$  over the class  $\Sigma$  defined by (1.2) in the model (1.1) satisfies:*

$$\mathcal{R}_{n,p}(\Sigma, \mu, x_0) \geq C(c, Q, p) r_n,$$

$$\text{where } C(c, Q, p) \triangleq \frac{c}{2^{1/p}} \left( e^{-Q} \vee \frac{1 - \sqrt{Q/2}}{2} \right)^{1/p}.$$

*Proof of theorem 3.7.* In view of proposition 5.4 all we have to do is to find two functions  $f_{0,n}$  and  $f_{1,n}$  such that:

- (1) There is some  $0 < Q < +\infty$  such that  $\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n) \leq Q$ ,
- (2)  $f_{0,n}, f_{1,n} \in \Sigma_{h_n, \alpha_n}(x_0, \omega)$ ,
- (3)  $|f_{0,n}(x_0) - f_{1,n}(x_0)| \geq 2cr_n$  for some constant  $c > 0$ .

We choose the 2 following hypothesis:

$$f_{0,n}(x) = \omega(h_n) \mathbf{1}_{|x-x_0| \leq h_n}, \quad f_{1,n}(x) = \omega(|x-x_0|) \mathbf{1}_{|x-x_0| \leq h_n},$$

(1): Since the  $\xi_i$  are centered Gaussian of variance  $\sigma^2$  and independants of  $\mathfrak{X}_n$  we have:

$$\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n | \mathfrak{X}_n) = \frac{1}{2\sigma^2} \sum_{i=1}^n (f_{0,n}(X_i) - f_{1,n}(X_i))^2,$$

thus by the definition 3.4 of  $h_n$ :  $\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n) = \frac{n}{2\sigma^2} \|f_{0,n} - f_{1,n}\|_{L^2(\mu)}^2 \leq n\omega^2(h_n) F_\nu(h_n) / \sigma^2 = 1$ .

(2): For  $h \in [0, h_n]$ , we have  $\omega_{f_{0,n},k}(x_0, h) = 0$  (take  $P$  as the constant polynomial equal

to  $\omega(h_n)$  and  $\omega_{f_{1,n,k}}(x_0, h) \leq \omega(h)$  (take  $P = 0$ ). Moreover for  $n$  large enough, we clearly have  $f_{0,n}, f_{1,n} \in \mathcal{U}(x_0, h_n, \alpha_n)$  since  $\alpha_n \rightarrow +\infty$ .

(3): If we take  $c = 1/2$  we have  $|f_{1,n}(x_0) - f_{0,n}(x_0)| = \omega(h_n) = 2cr_n$ .  $\square$

This lemma is used to compute the minimax risk of the examples in section 3.4.

**Lemma 5.5.** *Let  $\gamma > 0$  and  $\alpha \in \mathbb{R}$ . If  $G(h) = h^\gamma(\log(1/h))^\alpha$ , then we have:*

$$G^{\leftarrow}(h) \sim \gamma^{\alpha/\gamma} h^{1/\gamma} (\log(1/h))^{-\alpha/\gamma} \text{ as } h \rightarrow 0^+.$$

*Proof.* When  $\alpha = 0$ , the result is obvious. We assume  $\alpha \in \mathbb{R}^*$ . We look for  $h$  such that  $h^\gamma(\log(1/h))^\alpha = x$ , when  $x > 0$  is small. If  $\alpha > 0$  we define  $t = \log(h^{\gamma/\alpha})$ , so this equation becomes

$$(5.7) \quad t \exp(t) = -\gamma x^{1/\alpha} / \alpha,$$

where  $t \leq 0$ . The equation (5.7) has two solutions for  $x$  small enough, but they cannot be written in an explicit way. Then let us consider the Lambert function  $W$  defined as the function satisfying  $W(z)e^{W(z)} = z$  for any  $z \in \mathbb{C}$ . See for instance Corless *et al.* (1996) about this function. We are only interested here by its real branches. This function has two branches  $W_0$  and  $W_{-1}$  in  $\mathbb{R}$ . We denote by  $W_0$  the one such that  $W_0(0) = 0$  and  $W_{-1}$  the one such that  $\lim_{h \rightarrow 0^-} W_{-1}(h) = -\infty$ . The two solutions of (5.7) are then  $t_0 = W_{-1}(-\gamma x^{1/\alpha} / \alpha)$  and  $t_1 = W_0(-\gamma x^{1/\alpha} / \alpha)$  and  $h_0 \triangleq \exp(\alpha W_{-1}(-\gamma x^{1/\alpha} / \alpha) / \gamma)$  is the smallest solution. By the definition of  $W$  we have for  $-1/e < x < 0$  and  $a \in \mathbb{R}$ :  $e^{aW_{-1}(x)} = (-x)^a (-W_{-1}(x))^{-a}$  and since  $W_{-1}$  satisfies  $W_{-1}(-x) \sim \log(x)$  as  $x \rightarrow 0^+$  we have  $h_0 = (\gamma x^{1/\alpha} / \alpha)^{\alpha/\gamma} (-W_{-1}(-\gamma x^{1/\alpha} / \alpha))^{-\alpha/\gamma} \sim \gamma^{\alpha/\gamma} x^{1/\alpha} (\log(1/x))^{-\alpha/\gamma}$  as  $x \rightarrow 0^+$ . When  $\alpha < 0$  we procede similarly. We have  $t \geq 0$  and (5.7) has a single solution  $t = W_0(-\gamma x^{1/\alpha} / \alpha)$ , thus  $h \triangleq \exp(-\alpha W_0(-\gamma x^{1/\alpha} / \alpha) / \gamma)$ . By the definition of  $W_0$  we have  $\forall x > 0$  and  $a \in \mathbb{R}$ :  $e^{aW_0(x)} = x^a W_0^{-a}(x)$  and since  $W_0$  satisfies  $W_0(x) \sim \log(x)$  as  $x \rightarrow +\infty$  we find again  $h \sim \gamma^{\alpha/\gamma} x^{1/\alpha} (\log(1/x))^{-\alpha/\gamma}$  as  $x \rightarrow 0^+$ .  $\square$

## 5.2. $\Gamma$ -varying design.

*Proof of proposition 4.3.* We prove inequality (4.6), since the proof of (4.5) is similar and more direct. We denote the weights of the Nadaraya-Watson estimator by:

$$w_{n,i}(x_0) \triangleq \frac{K_{\rho,h}(X_i - x_0)}{\bar{K}_{n,\rho,h}(X_i - x_0)}.$$

We have  $|\tilde{f}_n(x_0, h) - f(x_0)| \leq \sum_{i=1}^n |f(X_i) - f(x_0)| w_{n,i}(x_0) + |\sum_{i=1}^n \xi_i w_{n,i}(x_0)|$ . In view of assumption K and since  $f \in \mathcal{F}_{\delta_n}(x_0, \omega)$ , we have for  $h$  small enough  $|f(X_i) - f(x_0)| \leq \omega(h + \rho(h)) \leq \omega(2h) \leq 2\omega(h)$ , where we used  $\omega \in \text{RV}_0(s)$  with  $0 < s \leq 1$ , and  $\rho(h) \ll h$ , in view of lemma B.2. Then the result follows, since  $\sum_{i=1}^n \xi_i w_{n,i}(x_0)$  is, conditional on  $\mathfrak{X}_n$ , a centered Gaussian variable of variance  $\sigma^2 \sum_{i=1}^n w_{n,i}^2(x_0) \leq \sigma^2 \bar{K}_{n,\rho,h}^{-1}$ .  $\square$

*Proof of proposition 4.6.* Let us define  $G(h) \triangleq \omega^2(h) F_\nu(h)$ . Let  $y \in \mathbb{R}$ . In view of theorem A.1 and lemma B.2 we get  $\lim_{h \rightarrow 0^+} \ell_\omega(h + y\rho(h)) / \ell_\omega(h) = 1$ , thus  $\lim_{h \rightarrow 0^+} \omega(h + y\rho(h)) / \omega(h) = 1$ . Moreover  $F_\nu \in \Gamma V_0(\rho)$  in view of proposition B.4, then  $\lim_{h \rightarrow 0^+} G(h + y\rho(h)) / G(h) = \exp(y)$ , and then  $G \in \Gamma V_0(\rho)$ . The function  $G$  is continuous and such that  $\lim_{h \rightarrow 0^+} G(h) = 0$ , so for  $n$  large enough,  $h_n$  is well defined and given by  $h_n = G^{\leftarrow}(\sigma^2/n)$ . Then theorem B.7 entails  $G^{\leftarrow} \in \Pi V_0(\ell)$  for  $\ell = \rho \circ \nu^{\leftarrow}$ , with  $\ell$  slowly varying. Moreover, the lemma B.6 entails  $G^{\leftarrow}$  is in particular slowly varying. Then  $h_n = G^{\leftarrow}(\sigma^2/n) \sim G^{\leftarrow}(n^{-1})$  as  $n \rightarrow +\infty$ , and the proposition follows with the choice  $\varphi \triangleq G^{\leftarrow}$ .  $\square$

*Proof of proposition 4.4.* We prove inequality (4.8). Let us define  $Z_{i,n} = K_{\rho,\gamma_n}(X_i - x_0) - \mathbb{E}_\mu^n\{K_{\rho,\gamma_n}(X_i - x_0)\}$ . Assumptions M, K entail for  $n$  large enough:

$$\frac{1}{\rho(\gamma_n)\nu(\gamma_n)}\mathbb{E}_\mu^n\{K_{\rho,\gamma_n}(X_1 - x_0)\} = \frac{2}{\nu(\gamma_n)}\int K(y)(\nu(\gamma_n + y\rho(\gamma_n))dy.$$

In view of assumption G and proposition B.1 we get:

$$\lim_{n \rightarrow +\infty} \frac{1}{\nu(\gamma_n)}\int K(y)\nu(\gamma_n + y\rho(\gamma_n))dy = \int K(y)e^y dy,$$

then for  $n$  large enough one has in view of lemma B.3:

$$\left\{ \left| \frac{\overline{K}_{n,\rho,\gamma_n}}{2nF_\nu(h_n)} - c_K \right| > \varepsilon \right\} \subset \left\{ \left| \frac{1}{n\rho(\gamma_n)\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n} \right| > \varepsilon/2 \right\}.$$

The variables  $(Z_{i,n})_{i=1,\dots,n}$  are clearly independant and centered, and assumption K entails  $|Z_{i,n}| \leq 2$ . In the same way as previously one has for  $n$  large enough

$$b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}_\mu^n\{Z_{i,n}^2\} = n\mathbb{E}_\mu^n\{Z_{1,n}^2\} \leq 3nF_\nu(\gamma_n) \int K^2(y)e^y dy,$$

thus we can apply the Bernstein inequality: we define  $\tau_n \triangleq \frac{\varepsilon}{2}nF_\nu(\gamma_n)$ , and we get:

$$\mathbb{P}_\mu^n \left\{ \left| \frac{\overline{K}_{n,\rho,\gamma_n}}{2nF_\nu(\gamma_n)} - c_K \right| > \varepsilon \right\} \leq 2 \exp\left(\frac{-\tau_n^2}{2(b_n^2 + \tau_n/3)}\right) \leq 2 \exp(-C_1\varepsilon^2 nF_\nu(\gamma_n)),$$

thus the inequality (4.8). Introducing the variables  $Z_{i,n} \triangleq \mathbf{1}_{|X_i - x_0| \leq \gamma_n} - \mathbb{P}_\mu^n\{|X_i - x_0| \leq \gamma_n\}$  and using again proposition B.3 and the Bernstein inequality, we prove (4.7) in the same fashion as (4.8).  $\square$

*Proof of theorem 4.7.* We prove inequality (4.13). Let  $0 < \varepsilon \leq 1$ , and let define the event:

$$\mathcal{A}_{n,\varepsilon} = \left\{ \left| \frac{\overline{K}_{n,\rho,h_n}}{2F_\nu(h_n)} - c_K \right| \leq \varepsilon \right\}.$$

Since  $nF_\nu(h_n) \rightarrow +\infty$  and  $\eta_n \rightarrow 0$  as  $n \rightarrow +\infty$ , we have for  $n$  large enough  $\mathcal{A}_{n,\varepsilon} \subset \{\overline{K}_{n,\rho,h_n} > \eta_n\}$ . Thus for  $n$  large enough, proposition 4.3 and proposition 4.4 together entail, in view of the definition 4.5 of  $h_n$  that  $|\tilde{f}_n(x_0) - f(x_0)|\mathbf{1}_{\mathcal{A}_{n,\varepsilon}} \leq 2\omega(h_n)(1 + (c_K - \varepsilon)^{-1/2}|\gamma_{h_n}|)$ . Thus we have

$$\sup_{f \in \mathcal{F}_{h_n}(x_0, \omega)} \mathbb{E}_{f,\mu}^n \left\{ r_n^{-p} |\tilde{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}} | \mathfrak{X}_n \right\} \leq C(K, p),$$

where  $C(K, p)$  is a constant depending on  $K$  and  $p$  only. Now we work on the complementary event  $\mathcal{A}_{n,\varepsilon}^c$ . If  $\overline{K}_{n,\rho,h_n} = 0$ , since  $\tilde{f}_n(x_0) = 0$  and  $f \in \mathcal{U}(x_0, h_n, \alpha_n)$  we have in view of (4.8):

$$\begin{aligned} \mathbb{E}_{f,\mu}^n \left\{ r_n^{-p} |\tilde{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}^c} \mathbf{1}_{\overline{K}_{n,\rho,h_n}=0} \right\} &\leq r_n^{-p(1+\gamma)} \mathbb{P}_{f,\mu}^n \left\{ \mathcal{A}_{n,\varepsilon}^c \right\} \\ &\leq r_n^{-p(1+\gamma)} \exp(-C_1\sigma^2 r_n^{-2}) = o_n(1). \end{aligned}$$

When  $\overline{K}_{n,\rho,h_n} > \eta_n$ , proposition 4.3 entails for  $n$  large enough:

$$|\tilde{f}_n(x_0) - f(x_0)| \leq 2\omega(h_n) + \sigma \overline{K}_{n,\rho,h_n}^{-1/2} |\gamma_{h_n}| \leq 2\eta_n^{-1/2} (1 + \sigma |\gamma_{h_n}|),$$

thus

$$\begin{aligned} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\tilde{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}^c} \mathbf{1}_{\overline{K}_{n,\rho,h_n} > \eta_n} \} &\leq C(\sigma, p) r_n^{-p} \eta_n^{-p/2} \mathbb{P}_{f,\mu}^n \{ \mathcal{A}_{n,\varepsilon}^c \} \\ &\leq C(\sigma, p) r_n^{-p(1+\tau/2)} \exp(-C_1 \sigma^2 r_n^{-2}) = o_n(1), \end{aligned}$$

uniformly for  $f \in \Sigma_{h_n, \alpha_n}(x_0, \omega)$ . Now assumption K entails

$$\mathbb{E}_{f,\mu}^n \{ |\tilde{f}_n(x_0)|^p \mathbf{1}_{\overline{K}_{n,\rho,h_n} \leq \eta_n} | \mathfrak{X}_n \} \leq 2^p (\alpha_n + \eta_n^{-p+1/2} C(p)),$$

thus:

$$\sup_{f \in \mathcal{U}(x_0, h_n, \alpha_n)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\tilde{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}^c} \mathbf{1}_{\overline{K}_{n,\rho,h_n} \leq \eta_n} \} = o_n(1),$$

and (4.13) follows. One can prove in the same way (4.12), using (4.7) and (4.5). Since  $r_n = \omega(h_n) = \omega(\varphi(n^{-1}))$  where  $\omega$  is regularly varying and  $\varphi$  is slowly varying, we have in view of proposition 4.6 that defining  $\ell_{\omega,\nu} \triangleq \omega \circ \varphi$  we get  $r_n = \ell_{\omega,\nu}(n^{-1})$  with  $\ell_{\omega,\nu}$  slowly varying.  $\square$

*Proof of theorem 4.8.* We introduce the same hypothesis as in the proof of theorem 3.7, with  $h_n$  given by definition 4.5. The proof is then exactly the same.  $\square$

5.2.1. *Proof for the example in section 4.4.* The equation to be solved is:

$$(5.8) \quad h^{1+2s+\alpha} \exp(-1/h^\alpha) = y_n,$$

where  $y_n \triangleq \sigma^2 \alpha / (r^2 n)$ . Defining  $t \triangleq h^{-\alpha}$ , equation (5.8) becomes  $t^{-(1+2s+\alpha)/\alpha} \exp(-t) = y_n$  that we rewrite  $x \exp(x) = \alpha / (1 + 2s + \alpha) y_n^{-\alpha/(1+2s+\alpha)}$ , for  $x \triangleq \alpha / (1 + 2s + \alpha) t$ . Then we have  $x = W_0(\alpha / (1 + 2s + \alpha) y_n^{-\alpha/(1+2s+\alpha)})$ , where  $W_0$  is defined in the proof of the lemma 5.5. Using the fact that  $W_0(x) \sim \log(x)$  as  $x \rightarrow +\infty$ , we get  $x \sim \frac{\alpha}{1+2s+\alpha} \log n$  as  $n \rightarrow +\infty$ , thus  $h_n \sim (\log n)^{-1/\alpha}$  as  $n \rightarrow +\infty$ , and the result holds since  $r_n \triangleq r h_n^s$ .  $\square$

## APPENDIX A. REGULAR VARIATION : KARAMATA THEORY

We recall here some results about regularly varying functions. The results stated in this section can be found in Bingham *et al.* (1989), Geluk and de Haan (1987) and Senata (1976).

**Theorem A.1** (Uniform convergence). *If  $\ell$  is slowly varying at 0 then one has:*

$$\lim_{h \rightarrow 0^+} \ell(yh) / \ell(h) = 1,$$

uniformly on each compact set in  $(0, +\infty)$ .

The uniform convergence theorem is the most important result in the theory. It was first proved by Karamata in 1930.

*Remark.* It is of special importance to notice the uniformity does not hold on an interval including 0. This is the reason why some extra assumptions will be required in Abelian type theorems.

**Proposition A.2.** *If  $\ell_1$  and  $\ell_2$  varies slowly, one has:*

- (1)  $\ell_1^\alpha$  varies slowly for any  $\alpha \in \mathbb{R}$ ,
- (2)  $\ell_1 \times \ell_2$ ,  $\ell_1 + \ell_2$  vary slowly,
- (3) If  $\lim_{h \rightarrow 0^+} \ell_2(h) = 0$  then  $\ell_1 \circ \ell_2$  varies slowly.

**Proposition A.3.** *If  $R$  varies regularly with an index  $\gamma \neq 0$ , then as  $h \rightarrow 0^+$  one has*

$$R(h) \rightarrow \begin{cases} 0 & \text{if } \gamma > 0, \\ +\infty & \text{if } \gamma < 0. \end{cases}$$

The asymptotic behaviour of integrals of regularly varying functions plays a key role in the proofs of this paper. See the proof of lemma 5.1 for instance.

**Theorem A.4** (Karamata). *If  $\ell$  varies slowly, and if  $\gamma > -1$ , one has:*

$$(A.1) \quad \int_0^h t^\gamma \ell(t) dt \sim (1 + \gamma)^{-1} h^{1+\gamma} \ell(h) \text{ as } h \rightarrow 0^+,$$

and in particular  $h \mapsto \int_0^h t^\gamma \ell(t) dt$  is regularly varying of index  $1 + \gamma$ .

The following proposition is an extension to the Karamata theorem when  $\gamma = -1$ :

**Proposition A.5.** *If  $\ell$  varies slowly and  $\int_0^\eta \ell(t) dt/t < +\infty$  for some  $\eta > 0$ , then  $h \mapsto \int_0^h \ell(t) dt/t$  is slowly varying, and*

$$\lim_{h \rightarrow 0^+} \frac{1}{\ell(h)} \int_0^h \ell(t) \frac{dt}{t} = +\infty.$$

**Proposition A.6.** *If  $R$  is some positive monotone function such that  $h \mapsto \int_0^h R(t) dt$  belongs to  $\text{RV}_0(\gamma)$  for some  $\gamma \geq 0$  then  $R \in \text{RV}_0(\gamma - 1)$ .*

**Theorem A.7** (Abelian theorem). *If  $\ell$  is slowly varying at 0 and  $K$  is a function such that  $\int_0^1 t^{-\delta} K(t) dt < +\infty$  for some  $\delta > 0$  then:*

$$(A.2) \quad \int_0^1 K(t) \ell(th) dt \sim \ell(h) \int_0^1 K(t) dt \text{ as } h \rightarrow 0^+.$$

**Proposition A.8.** *If  $\ell$  varies slowly with  $\int_0^\eta \ell(t) dt/t < +\infty$  for some  $\eta > 0$ , and  $K$  is such that  $\forall t \in \mathbb{R}^+$ ,  $|K(t) - K(0)| \leq \rho |t|^\kappa$  for some  $\rho > 0$  and  $\kappa > 0$  one has:*

$$\int_0^1 K(t/h) \ell(t) dt/t \sim K(0) \int_0^1 \ell(t) dt/t \text{ as } h \rightarrow 0^+.$$

Now we give a result concerning the inversion of regularly varying functions. If  $R$  is defined and bounded on  $[0, +\infty[$  one can define the generalised inverse:

$$(A.3) \quad R^\leftarrow(y) = \inf\{h > 0 \text{ such that } R(h) \geq y\}.$$

**Theorem A.9** (Inversion theorem). *If  $R \in \text{RV}_0(\gamma)$  for some  $\gamma > 0$ , then there exists  $R^- \in \text{RV}_0(1/\gamma)$  such that*

$$(A.4) \quad R(R^-(h)) \sim R^-(R(h)) \sim h \text{ as } h \rightarrow 0^+,$$

and  $R^-$  is unique up to an asymptotic equivalence. Moreover, one version of  $R^-$  is  $R^{\leftarrow}$ .

**Theorem A.10.** *If  $(\delta_n)_{n \geq 0}$  and  $(\lambda_n)_{n \geq 0}$  are sequences of positive numbers such that  $\delta_{n+1} \sim \delta_n$  as  $n \rightarrow +\infty$ ,  $\lim_n \delta_n = 0$ , and if there is a positive and continuous function  $\phi$  such that for any  $y > 0$ :*

$$\lim_n \lambda_n R(y \delta_n) = \phi(y),$$

then  $R$  varies regularly.

APPENDIX B.  $\Gamma$ -VARIATION : DE HAAN THEORY

We describe here the properties of  $\Gamma$ -varying functions and  $\Pi$ -varying functions. The results are due to de Haan. The references are the same as for regular variation. All the following results can be found there in.

**Proposition B.1** (Uniformity). *If  $\nu$  is a function such that (4.1) holds for all  $y \in \mathbb{R}$ , then (4.1) holds uniformly on each compact set in  $\mathbb{R}$ .*

**Lemma B.2.** *If  $\rho$  is such that (4.1) holds, then:*

$$(B.1) \quad \lim_{h \rightarrow 0^+} \rho(h)/h = 0.$$

**Lemma B.3.** *The auxiliary function  $\rho$  in definition (4.1) is unique up to within an asymptotic equivalence and can be taken as  $h \mapsto \int_0^h \nu(t)dt/\nu(h)$ .*

**Proposition B.4.** *The class  $\Gamma V_0(\rho)$  is stable under integration. If  $\nu \in \Gamma V_0(\rho)$ , then  $F_\nu(h) = \int_0^h \nu(t)dt \in \Gamma V_0(\rho)$ .*

We have seen that under the operation of functional inversion, the class of regularly varying functions  $RV_0$  is stable. In the case of  $\Gamma$ -variation, the inversion maps the class  $\Gamma V_0$  in another class of functions, namely the de Haan class  $\Pi V_0$ .

**Definition B.5** ( $\Pi$ -Variation). A function  $\nu$  is in the de Haan class  $\Pi V_0$  if there exists a slowly varying function  $\ell$  and a positive real number  $c$  such that:

$$(B.2) \quad \forall y > 0, \quad \lim_{h \rightarrow 0^+} (\nu(yh) - \nu(h))/\ell(y) = c \log(y).$$

The class of functions  $\nu$  satisfying (B.2) is denoted by  $\Pi V_0(\ell)$ . The function  $\ell$  is called the *auxiliary* function of  $\nu$ .

The following lemma tells us  $\Pi$ -variation can be viewed as a refinement of slow variation: actually, any  $\Pi$ -varying function is slowly varying.

**Lemma B.6.** *For any  $\ell \in RV_0(0)$  we have:*

$$\Pi V_0(\ell) \subset RV_0(0).$$

**Theorem B.7.** *If  $\nu \in \Gamma V_0(\rho)$  then  $\ell = \rho \circ \nu^\leftarrow$  is slowly varying and  $\nu^\leftarrow \in \Pi V_0(\ell)$ . Conversely if  $\nu \in \Pi V_0(\ell)$  for some  $\ell \in RV_0(0)$  then  $\nu^\leftarrow \in \Gamma V_0(\rho)$  with  $\rho = \ell \circ \nu^\leftarrow$ . In both senses the inverses and their auxiliary functions are asymptotically unique.*

## REFERENCES

- [1] BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. *Regular Variation*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1989.
- [2] CAI, T., AND BROWN, L. Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics* 26 (1998), 1783–1799.
- [3] CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Society* 74 (1979), 829–836.
- [4] CORLESS, R. M., GONNET, G. H., HARE, D. E. G., AND JEFFREY, D. J. On the Lambert W function. *Advances in Computational Mathematics* 5 (1996), 329–359.
- [5] FAN, J., AND GILBELS, I. *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1996.
- [6] GELUK, J. L., AND DE HAAN, L. *Regular Variations, Extensions and Tauberian Theorems*. CWI Tract, 1987.
- [7] GOLDENSHLUGER, A., AND NEMIROVSKI, A. On spatially adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics* 6, 2 (1997), 135–170.



- [8] GUERRE, E. Efficient random rates for nonparametric regression under arbitrary designs. Private communication, dec 1999.
- [9] GUERRE, E. Design adaptive nearest neighbor regression estimation. *Journal of Multivariate Analysis* 75 (2000), 219–244.
- [10] HALL, P., MARRON, J. S., NEUMANN, M. H., AND TETTERINGTON, D. M. Curve estimation when the design density is low. *The Annals of Statistics* 25, 2 (1997), 756–770.
- [11] IBRAGIMOV, I., AND HASMINSKI, R. *Statistical Estimation: Asymptotic Theory*. Springer, New York, 1981.
- [12] KOROSTELEV, V., AND TSYBAKOV, A. *Minimax theory of image reconstruction*. Springer-Verlag, New York, 1993.
- [13] LEPSKI, O. V., MAMMEN, E., AND SPOKOINY, V. G. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics* 25, 3 (1997), 929–947.
- [14] LEPSKI, O. V., AND SPOKOINY, V. G. Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics* 25, 6 (1997), 2512–2546.
- [15] LEPSKI, O. V. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications* 35, 3 (1990), 454–466.
- [16] NEMIROVSKI, A. *Topics in Non-Parametric Statistics*. Ecole d’été de probabilités de Saint-Flour XXVIII - 1998. Lecture Notes in Mathematics, no. 1738. Springer, New York, 2000.
- [17] RESNICK, S. I. *Extreme Values, Regular Variation and Point Processes*. Applied Probability. Springer-Verlag, 1987.
- [18] SENATA, E. *Regularly Varying Functions*. Lecture Notes in Mathematics. Springer-Verlag, 1976.
- [19] SPOKOINY, V. G. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics* 26, 4 (1998), 1356–1378.
- [20] STONE, C. J. Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* 8 (1980), 1348–1360.
- [21] STONE, C. J. Consistent nonparametric regression. *The Annals of Statistics* 5 (1985), 595–645.
- [22] TSYBAKOV, A. Robust reconstruction of functions by the local approximation. *Problems of Information Transmission* 22 (1986), 133–146.
- [23] TSYBAKOV, A. *Introduction à l’estimation non-paramétrique*. Springer, 2003.

LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES, U.M.R. CNRS 7599 AND UNIVERSITÉ PARIS 7, 175 RUE DU CHEVALERET, 75013 PARIS  
*E-mail address:* gaiffas@math.jussieu.fr