Volume 14, No. 1 (2005), pp. Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

CONVERGENCE RATES FOR POINTWISE CURVE ESTIMATION WITH A DEGENERATE DESIGN

S. Gaïffas

Labor. Probab. et Modèles Aléatoires, U.M.R. CNRS 7599 and Univ. Paris 7
175 rue du Chevaleret, 75013 Paris
E-mail: gaiffas@math.jussieu.fr

The nonparametric regression with a random design model is considered. We want to recover the regression function at a point x_0 where the design density is vanishing or exploding. Depending on assumptions on 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 $\ell(n)$ rates, where ℓ is slowly varying (for instance $(\log n)^{-1}$), to fast $n^{-1/2}\ell(n)$ rates. If the continuity modulus of the regression function at x_0 can be bounded from above by an s-regularly varying function, and if the design density is β-regularly varying, we prove that the minimax convergence rate at x_0 is $n^{-s/(1+2s+\beta)}\ell(n)$.

Key words: degenerate design, minimax, nonparametric regression, random design

2000 Mathematics Subject Classification: 62G05, 62G08.

1. Introduction

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

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where $f\colon \mathbb{R} \to \mathbb{R}$, the variables (ξ_i) are centered Gaussian of variance σ^2 and independent of $X_1, \ldots, X_n$ (the design), and the X_i are distributed with density μ. We want to recover f at a chosen x_0.

For instance, if we take the variables (X_i) distributed with 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 $\beta > -1$, then clearly when $\beta > 0$ this density models a lack of information at x_0 and conversely an exploding amount of information if $-1 < \beta < 0$. We want to understand the influence of the parameter β on the amount of information at x_0 in the minimax setup.

1.2. Motivations. The pointwise estimation of the 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 were Ibragimov and Hasminski (1981) and Stone (1977). Over the class of Hölder functions with smoothness s, the local polynomial estimator converges with the rate $n^{-s/(1+2s)}$ (see Stone (1977)) and this rate is 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 the design density to be non-vanishing and finite at the estimation point. This assumption roughly means that the information is spatially *homogeneous*. The next logical step is to look for the minimax risk at a point where the design density μ is vanishing or exploding. To achieve such a result, it seems natural to consider several types of design density behaviour at x_0 and to compute the corresponding minimax rates. Such results would improve the statistical description of models (here in the minimax setup) with very inhomogeneous information.

When f has a Hölder type smoothness of order 2 and if $\mu(x) \sim x^\beta$ near 0, where $\beta > 0$, Hall *et al.* (1997) show that a local linear procedure converges with the rate $n^{-4/(5+\beta)}$ when estimating f at 0. This rate is also proved to be optimal. In a more general setup for the design and if the regression function is Lipschitz, Guerre (1999) extends the result of Hall *et al.* for $\beta > -1$. Here, we intend to develop the regression function estimation for degenerate designs in a systematic way.

1.3. Organization of the paper. In Section 2 we present two theorems giving the pointwise minimax convergence rate in the model (1.1) for different design behaviours (Theorems 1 and 2). In Section 3 we construct an estimator and in Section 4 give upper bounds for this estimator (Propositions 4 and 5). In Section 5 we discuss some technical points. The proofs are delayed until Section 6 and well-known facts about the regular and Γ-variation are given in the Appendix.

2. Main Results

All along this study we are in the minimax setup. We define the pointwise minimax risk over a class Σ by

$$\mathcal{R}_n(\Sigma,\mu) \triangleq \Big(\inf_{T_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\{|T_n(x_0) - f(x_0)|^p\}\Big)^{1/p}, \tag{2.1}$$

where $\inf_{T_n}$ is taken over all estimators T_n based on the observations (1.1), with x_0 being the estimation point and $p > 0$. The expectation $\mathbb{E}^n_{f,\mu}$ in (2.1) is taken with respect to the joint probability distribution $\mathbb{P}^n_{f,\mu}$ of the pairs $(X_i, Y_i)_{i=1,\dots,n}$.

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

Definition 1 (Regular variation). A continuous function $\nu\colon \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at 0 if there is a real number $\beta \in \mathbb{R}$ such that:

$$\forall y > 0, \quad \lim_{h\to 0^+} \nu(yh)/\nu(h) = y^\beta. \tag{2.2}$$

We denote by $\mathrm{RV}(\beta)$ the set of all the functions satisfying (2.2). A function in $\mathrm{RV}(0)$ is *slowly varying.*

Remark. Roughly, a regularly varying function behaves as a power function times a slower term. Typical examples of such functions are x^β, $x^\beta(\log(1/x))^\gamma$ for $\gamma \in \mathbb{R}$, and more generally any power function times a log or a composition of log-functions to some power. For other examples, see the references cited above.

2.2. The functions class

Definition 2. If $\delta > 0$ and $\omega \in \mathrm{RV}(s)$ with $s > 0$ we define the class $\mathcal{F}_\delta(x_0, \omega)$ of functions $f\colon [0,1] \to \mathbb{R}$ such that

$$\forall h \le \delta, \quad \inf_{P\in\mathcal{P}_k} \sup_{|x-x_0|\le h} |f(x) - P(x-x_0)| \le \omega(h),$$

where $k = \lfloor s \rfloor$ (the largest integer smaller than s) and $\mathcal{P}_k$ is the set of all the real polynomials with degree k. We define $\ell_\omega(h) \triangleq \omega(h)h^{-s}$, the slow variation term of ω. If $\alpha > 0$, we define

$$\mathcal{U}(\alpha) \triangleq \big\{ f\colon [0,1] \to \mathbb{R} \text{ such that } \|f\|_\infty \le \alpha \big\}.$$

Finally, we define

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

Remark. If we take $\omega(h) = rh^s$ for some $r > 0$, then we get the classical Hölder regularity with radius r. In this sense, the class $\mathcal{F}_\delta(x_0, \omega)$ is a slight generalization of the Hölder regularity.

Assumption M. In what follows, we assume that there exists a neighbourhood W of x_0 and a continuous function $\nu\colon \mathbb{R}^+ \to \mathbb{R}^+$ such that:

$$\forall x \in W, \quad \mu(x) = \nu(|x - x_0|). \tag{2.3}$$

This assumption roughly means that close to x_0 there are as many observations on the left of x_0 as on the right. All the following results can be extended easily to the non-symmetric case, see Section 5.1.

2.3. Regularly varying design density.

Theorem 1 gives the minimax rate over the class Σ (see Definition 2) for the estimation problem of f at x_0 when the design is regularly varying at this point.

We denote by $\mathcal{R}(x_0, \beta)$ the set of all the densities μ such that (2.3) holds with $\nu \in \mathrm{RV}(\beta)$ for a fixed neighbourhood W.

Theorem 1. *If*

- $(s, \beta) \in (0, +\infty) \times (-1, +\infty)$ *or* $(s, \beta) \in (0, 1] \times \{-1\}$*,*
- $\Sigma = \Sigma_{h_n,\alpha_n}(x_0, \omega)$ *with* $\omega \in \mathrm{RV}(s)$*,* $\alpha_n = O(n^\gamma)$ *for some* $\gamma > 0$ *and* h_n *given by* (2.5)*,*
- $\mu \in \mathcal{R}(x_0, \beta)$*,*

then we have

$$\mathcal{R}_n(\Sigma,\mu) \asymp \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(n^{-1}) \qquad \text{as} \quad n \to +\infty, \tag{2.4}$$

where $\ell_{\omega,\nu}$ *is slowly varying and where* $\asymp$ *stands for the equality in order, up to constants depending on* s*,* β *and* p *(see (2.1)) but not on* σ*. Moreover, the minimax rate is equal to* $\omega(h_n)$*, where* h_n *is the smallest solution to*

$$\omega(h) = \frac{\sigma}{\sqrt{2n \int_0^h \nu(t)\,dt}}. \tag{2.5}$$

Example. The simplest example is the non-degenerate design case ($0 < \mu(x_0) < +\infty$) with the class Σ equal to a Hölder ball ($\omega(h) = rh^s$, see Definition 2). This is the common case found in the literature. In particular, in this case, the design is slowly varying ($\beta = 0$ with the slow term constant and equal to $\lim_{x \to x_0} \mu(x)$). Solving (2.5) leads to the classical minimax rate

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

Example. Let $\beta > -1$. We consider ν such that $\int_0^h \nu(t)\,dt = h^{\beta+1}(\log(1/h))^\alpha$ and $\omega(h) = rh^s(\log(1/h))^\gamma$, where α, γ are any real numbers. In this case, we find that the minimax rate (see Section 6.5 for details) is

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

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

Example. Let $\beta = -1$, $\alpha > 1$, and $\nu(h) = h^{-1}(\log(1/h))^{-\alpha}$. Let ω be the same as in the previous example with $0 < s \le 1$. Then the minimax convergence rate is

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

This rate is almost the parametric estimation rate, up to the slow log factor. This result is natural since the design is very "exploding": we have a lot of information at x_0, thus we can estimate $f(x_0)$ very fast. Also, we note that the regularity parameters of the regression function (r, s, and γ) have (asymptotically) disappeared from the minimax rate.

2.4. Γ-varying design density. The regular variation framework includes any design density behaving close to the estimation point as a polynomial times a slow term. It does not include, for instance, a design with a behaviour similar to $\exp(-1/|x - x_0|)$ and defined as 0 at x_0, since this function goes to 0 at x_0 faster than any power function.

Such a local behaviour can model the situation where we have very little information. This example naturally leads us to the framework of Γ-variation. In fact, such a function belongs to the following class introduced by de Haan (1970).

Definition 3 (Γ-variation). A non-decreasing continuous function $\nu\colon \mathbb{R}^+ \to \mathbb{R}^+$ is Γ-varying if there exists a continuous function $\rho\colon \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\forall y \in \mathbb{R}, \quad \lim_{h\to 0^+} \nu(h + y\rho(h))/\nu(h) = \exp(y). \tag{2.6}$$

We denote by $\Gamma\mathrm{V}(\rho)$ the class of all such functions. The function ρ is called the *auxiliary* function of ν.

Remark. A function behaving like $\exp(-1/|x - x_0|)$ close to x_0 satisfies Assumption M with $\nu(h) = \exp(-1/h)$, where $\nu \in \Gamma\mathrm{V}(\rho)$ with $\rho(h) = h^2$.

Theorem 2. *If*

- $\Sigma = \Sigma_{h_n,\alpha_n}(x_0, \omega)$, *where* $\omega \in \mathrm{RV}(s)$ *with* $0 < s \leq 1$, h_n *is given by* (2.5) *and* $\alpha_n = O(r_n^{-\gamma})$ *for some* $\gamma > 0$ *with* $r_n \triangleq \omega(h_n)$,
- μ *satisfies Assumption* M *with* $\nu \in \Gamma\mathrm{V}(\rho)$,

then

$$\mathcal{R}_n(\Sigma, \mu) \asymp \ell_{\omega,\nu}(n^{-1}) \qquad \text{as} \quad n \to +\infty,$$

where $\ell_{\omega,\nu}$ *is slowly varying. Moreover, as in Theorem* 1, *the minimax rate is equal to* $\omega(h_n)$, *where* h_n *is the smallest solution to* (2.5).

Example. Let μ satisfy Assumption M with $\nu(h) = \exp(-1/h^\alpha)$ for $\alpha > 0$ and $\omega(h) = rh^s$ for $0 < s \leq 1$. It is an easy computation to see that ν belongs to the class $\Gamma\mathrm{V}(\rho)$ for the auxiliary function $\rho(h) = \alpha^{-1}h^{\alpha+1}$. In this case, we find that the minimax rate (see Section 6.5 for details) is

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

As shown by Theorem 2, we find a very slow minimax rate in this example. We note that the parameters s and α are on the same scale.

3. Local Polynomial Estimation

3.1. Introduction. For the proof of the upper bound in Theorem 1 we use a local polynomial estimator. The local polynomial estimator is well-known and has been intensively studied (see Stone (1977), Fan and Gijbels (1996), Spokoiny (1998), Tsybakov (2003), among many others). If f is a smooth function at x_0, then it is close to its Taylor polynomial. 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

$$f(x) \approx f(x_0) + f^{'}(x_0)(x - x_0) + \ldots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k. \tag{3.1}$$

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

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

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

$$\langle f, g\rangle_{h,K} \triangleq \sum_{i=1}^{n} f(X_i)g(X_i)K\Big(\frac{X_i - x_0}{h}\Big), \tag{3.2}$$

and the corresponding pseudo-norm $\|\cdot\|_{h,K} \triangleq \sqrt{\langle \cdot, \cdot\rangle_{h,K}}$ ($K \geq 0$). In view of (3.1) it is natural to consider the estimator defined as the closest polynomial of degree k to the observations (Y_i) in the least square sense, that is:

$$\widehat{f}_h = \operatorname*{argmin}_{g \in V_{k,h}} \|g - Y\|_{h,K}^2. \tag{3.3}$$

Then $\widehat{f}_h(x_0)$ is the *local polynomial estimator* of f at x_0. A necessary condition for $\widehat{f}_h$ to be the minimizer of (3.3) is that it solves the linear problem:

$$\text{find } \widehat{f} \in V_{k,h} \text{ such that } \forall\, \phi \in V_{k,h}, \quad \langle \widehat{f}, \phi\rangle_{h,K} = \langle Y, \phi\rangle_{h,K}. \tag{3.4}$$

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

$$\widehat{f}_h = P_{\widehat{\theta}_h}, \tag{3.5}$$

where

$$P_\theta = \theta_0\phi_{0,h} + \theta_1\phi_{1,h} + \ldots + \theta_k\phi_{k,h}, \tag{3.6}$$

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

$$\mathbf{X}_h^K \theta = \mathbf{Y}_h^K, \tag{3.7}$$

where $\mathbf{X}_h^K$ is the symmetric matrix with entries

$$(\mathbf{X}_h^K)_{j,l} = \langle \phi_{j,h}, \phi_{l,h}\rangle_{h,K}, \quad 0 \leq j, l \leq k, \tag{3.8}$$

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. Let K be the rectangular kernel $K^R(x) = \frac{1}{2}\mathbf{1}_{|x|\leq 1}$ or a non-negative function such that:

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

Assumption K is satisfied by all the classical kernels used in nonparametric curve smoothing. Let us define

$$N_{n,h} = \#\{X_i \text{ such that } X_i \in [x_0 - h, x_0 + h]\}, \tag{3.9}$$

the number of observations in the interval $[x_0 - h, x_0 + h]$, and the random matrix

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

Denote $\mathfrak{X}_n \triangleq \sigma(X_1, \ldots, X_n)$ the σ-algebra generated by the design. Note that $\mathcal{X}_h^K$ is measurable with respect to $\mathfrak{X}_n$. The matrix $\mathcal{X}_h^K$ is a "renormalization" of $\mathbf{X}_h^K$. We show in Lemma 6 that this matrix is asymptotically non-degenerate with large probability when the design is regularly varying.

For technical reasons, we introduce a slightly different version of the local polynomial estimator. We introduce a "correction" term in the matrix $\mathbf{X}_h^K$.

Definition 4. Given some $h > 0$, we consider $\widehat{f}_h$ defined by (3.5) 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

$$\widetilde{\mathbf{X}}_h^K \theta = \mathbf{Y}_h^K, \tag{3.10}$$

where

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

with $\lambda(M)$ being 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 $\widetilde{\mathbf{X}}_h^K$ as follows: in the "good" case when $\mathcal{X}_h^K$ is non-degenerate in the sense that its smallest eigenvalue is not too small, we solve the system (3.7), while in the "bad" case we still have a control on the smallest eigenvalue of $\widetilde{\mathbf{X}}_h^K$, since we always have $\lambda(\widetilde{\mathbf{X}}_h^K) \ge N_{n,h}^{1/2}$.

3.2. Bias-variance equilibrium. A main result on the local polynomial estimator is the bias-variance decomposition. This is a classical result presented many times in different forms: see Cleveland (1979), Goldenshluger and Nemirovski (1997), Korostelev and Tsybakov (1993), Spokoiny (1998), Stone (1980), Tsybakov (1986, 2003). The version in Spokoiny (1998) is close to the one presented here. The differences are mostly related to the fact that the design is random and that we consider a modified version of the local polynomial estimator (see Definition 4). We introduce the event

$$\Omega_h^K \triangleq \{X_1, \ldots, X_n \text{ are such that } \lambda(\mathcal{X}_h^K) > N_{n,h}^{-1/2} \text{ and } N_{n,h} > 0\}. \tag{3.11}$$

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

Proposition 1 (Bias–variance decomposition). *Under Assumption* K *and if* $f \in \mathcal{F}_h(x_0, \omega)$*, the following inequality holds on the event* Ω_h^K:

$$|\widehat{f}_h(x_0) - f(x_0)| \le \lambda^{-1}(\mathcal{X}_h^K)\sqrt{k+1} K_\infty \big(\omega(h) + \sigma N_{n,h}^{-1/2} |\gamma_h|\big), \tag{3.12}$$

where γ_h *is, conditionally on* $\mathfrak{X}_n$*, centered Gaussian such that* $\mathbb{E}_{f,\mu}^n\{\gamma_h^2 \mid \mathfrak{X}_n\} \le 1$.

Remark. Inequality (3.12) holds conditionally on the design, on the event Ω_h^K. We will see that this event has a large probability in the regular variation framework.

3.3. Choice of the bandwidth. Now, like with any linear estimation procedure, the problem is: *how to choose the bandwidth* h? In view of inequality (3.12) a natural bandwidth choice is

$$H_n \triangleq \operatorname*{argmin}_{h \in [0,1]} \Big\{ \omega(h) \geq \frac{\sigma}{\sqrt{N_{n,h}}} \Big\}. \tag{3.13}$$

Such a bandwidth choice is well known, see, for instance, Guerre (2000). This choice stabilizes the procedure, since it is sensitive to the design, which represents in the model (1.1) the local amount of information. The estimator is then defined by

$$\widehat{f}_n(x_0) \triangleq \widehat{f}_{H_n}(x_0),$$

where $\widehat{f}_h$ is given by Definition 4 and H_n is defined by (3.13). The random bandwidth H_n is close in probability to the theoretical deterministic bandwidth h_n defined by (2.5) in view of the following proposition.

Proposition 2. *Under Assumption* M *and if* $\omega \in \mathrm{RV}(s)$ *for any* $s > 0$*, for any* $0 < \varepsilon \leq 1/2$ *there exists* $0 < \eta \leq \varepsilon$ *such that*

$$\mathbb{P}_\mu^n \Big\{ \Big| \frac{H_n}{h_n} - 1 \Big| > \varepsilon \Big\} \leq 4 \exp\Big(- \frac{\eta^2}{1 + \eta/3} n F_\nu(h_n/2) \Big),$$

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

If $nF_\nu(h_n/2) \to +\infty$ as $n \to +\infty$ (this is the case when ν is regularly varying) this inequality entails

$$H_n = \big(1 + o_{\mathbb{P}_{f,\mu}^n}(1)\big) h_n,$$

where $o_{\mathbb{P}}(1)$ stands for a sequence going to 0 in probability under a probability $\mathbb{P}$.

Proposition 3 motivates the regularly varying design choice. It makes a link between the behaviour of the counting process $N_{n,h}$ (that appears in the variance term of (3.12)) and the behaviour of μ close to x_0. Actually, the regular variation property (see Definition 1) naturally appears under appropriate assumptions on the asymptotic behaviour of $N_{n,h}$. Let us denote by $\mathbb{P}_\mu^n$ the joint probability of the variables (X_i).

Proposition 3. *If Assumption* M *holds with* ν *monotone, then the following properties are equivalent*:

(1) ν *is regularly varying of index* $\beta \geq -1$;

(2) *there exist sequences of positive numbers* (λ_n) *and* (γ_n) *such that* $\lim_n \gamma_n = 0$, $\liminf_n n\lambda_n^{-1} > 0$, $\gamma_{n+1} \sim \gamma_n$ *as* $n \to +\infty$ *and a continuous function* $\phi\colon \mathbb{R}^+ \to \mathbb{R}^+$ *such that for any* $C > 0$:

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

(3) *there exist* (λ_n), (γ_n), *and* ϕ *as before such that for any* $C > 0$ *and* $\varepsilon > 0$:

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

The proof is delayed until Section 6. Mainly, it is a consequence of the sequence characterization of regular variation (see in the Appendix).

4. Upper Bounds for $\widehat{f}_{H_n}(x_0)$

4.1. Conditional on the design. When no assumptions on the behavior of the design density are made, we can work conditionally on the design. For $\lambda > 0$ we define the event

$$\mathrm{E}_\lambda \triangleq \{\lambda_n > \lambda\},$$

where $\lambda_n \triangleq \lambda(\mathcal{X}_{H_n}^K)$. Note that $\mathrm{E}_\lambda \in \mathfrak{X}_n$. We also define the constant

$$m(p) \triangleq \sqrt{2/\pi} \int_{\mathbb{R}^+} (1+t)^p \exp(-t^2/2)\, dt.$$

Proposition 4. *Under Assumption* K, *if* λ *is such that* $\lambda^2 N_{n,H_n} \geq 1$ *and* $n \geq k+1$, *we have on* E_λ:

$$\sup_{f \in \mathcal{F}_{H_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_n(x_0) - f(x_0)|^p \mid \mathfrak{X}_n \big\} \leq m(p)\lambda^{-p} K_\infty^p (k+1)^{p/2} R_n^p,$$

where $R_n \triangleq \omega(H_n)$.

4.2. When the design is regularly varying. Proposition 5 below gives an upper bound for the estimator $\widehat{f}_{H_n}(x_0)$ when the design density is regularly varying. This proposition can be viewed as a deterministic counterpart to Proposition 4.

Let $\lambda_{\beta,K}$ be the smallest eigenvalue of the symmetric and positive matrix with entries, for $0 \leq j, l \leq k$:

$$(\mathcal{X}_{\beta,K})_{j,l} = \frac{\beta+1}{2} \big(1 + (-1)^{j+l}\big) \int_0^1 y^{j+l+\beta} K(y)\, dy. \tag{4.1}$$

Note that in view of Lemma 6 we have $\lambda_{\beta,K} > 0$.

Proposition 5. *Let* $\varrho > 1$ *and let* h_n *be defined by* (2.5). *Let* (α_n) *be a sequence of positive numbers such that* $\alpha_n = O(n^\gamma)$ *for some* $\gamma > 0$. *If* $\mu \in \mathcal{R}(x_0, \beta)$ *with* $\beta > -1$ *and* $\omega \in \mathrm{RV}(s)$, *we have for any* $p > 0$:

$$\limsup_n \sup_{f \in \Sigma_{\varrho h_n, \alpha_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \} \leq C \lambda_{\beta,K}^{-p}, \tag{4.2}$$

where $r_n \triangleq \omega(h_n)$ *satisfies*

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

with $\ell_{\omega,\nu}$ *slowly varying and where* $C = 4^{s/(1+2s+\beta)} (k+1)^{p/2} m(p) K_\infty^p$.

Remark. Under Hölder regularity with radius r we have

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

5. Discussion

5.1. About Assumption M. As stated previously, Assumption M means that the design distribution is symmetric around x_0 close to this point. When it is not the case, and if there are two functions $\nu^- \in \mathrm{RV}(\beta^-)$, $\nu^+ \in \mathrm{RV}(\beta^+)$ for $\beta^-, \beta^+ \geq -1$ and $\eta^-, \eta^+ > 0$ such that for any $x \in [x_0 - \eta^-, x_0 + \eta^+]$:

$$\mu(x) = \nu^+(x - x_0)\mathbf{1}_{x_0 \leq x \leq x_0 + \eta^+} + \nu^-(x_0 - x)\mathbf{1}_{x_0 - \eta^- \leq x < x_0},$$

we can easily prove that the minimax convergence rate is the fastest among the two possible ones, which is (2.4) for the choice of $\beta = \beta^- \wedge \beta^+$. To prove the upper bound we can use the same estimator as in Section 3 with a non-symmetric choice of the bandwidth, or more roughly we can "throw away" the observations on the side of x_0 corresponding to the largest index of regular variation (when μ is known).

5.2. On Theorem 1 and Propositions 4 and 5. Since we are interested in the estimation of f at x_0, we need only a regularity assumption in some neighbourhood of this point. Note that the minimax risks are computed over a class where the regularity assumption holds in a decreasing interval as n increases.

It appears that a natural choice of the size of this interval is the theoretical bandwidth of estimation h_n, since it is the minimum we need for the proof of the upper bounds. To state an upper bound with the "design-adaptive" estimator $\widehat{f}_{H_n}(x_0)$ — in the sense that it does not depend on the behavior of the design density close to x_0 (via the parameter β for instance) — we need a smoothness control in a slightly larger neighbourhood size than h_n (see the parameter ϱ in Proposition 5).

More precisely, to prove in Proposition 5 that r_n is an upper bound, we use, in particular, Proposition 2 with $\varepsilon = \varrho - 1$ in order to control the random bandwidth H_n by h_n. Thus, the parameter ϱ is indispensable for the proof of Proposition 5. Note that we do not need such a parameter in Theorem 1 since we use the estimator with the deterministic bandwidth h_n to prove the upper bound part of the theorem. Of course, this estimator in unfeasible from a practical point of view since h_n heavily depends on μ, which is hardly known in practice. This is the reason why we state Proposition 5, which tells us that the estimator with the data-driven bandwidth H_n converges with the same rate.

5.3. On Theorem 2. In the Γ-variation framework, for the proof of the upper bound part of Theorem 2 we use an estimator depending on μ. Again, such an estimator is unfeasible from a practical point of view. Anyway, this framework is considered only for theoretical purposes, since from a practical point of view nothing can be done in this case: there is no observations at the point of estimation. This is precisely what Theorem 2 and the corresponding example tell us, in the sense that the minimax rate is very slow.

5.4. About the Γ-varying design case. For the proof of the upper bound part in Theorem 2 we can consider an estimator different from the classical regressogram (see the proof of the theorem). If K is a kernel satisfying Assumption K, we define

$$\widetilde{f}_n(x_0) \triangleq \frac{\sum_{i=1}^n Y_i\big(K\big(\frac{X_i-h_n-x_0}{\rho(h_n)}\big)+K\big(\frac{X_i+h_n-x_0}{\rho(h_n)}\big)\big)}{\sum_{i=1}^n K\big(\frac{X_i-h_n-x_0}{\rho(h_n)}\big)+K\big(\frac{X_i+h_n-x_0}{\rho(h_n)}\big)},$$

where h_n is defined by (2.5). The point is that since $\operatorname{Supp} K \subset [-1,1]$, this estimator 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 does not contain the point of estimation x_0 for n large enough, since $\lim_{h\to 0^+} \rho(h)/h = 0$ (see Appendix). In spite of this, we can prove that $\widetilde{f}_n(x_0)$ converges with the rate r_n. We can understand this as follows: since there is no information at x_0, the procedure actually "catches" the information "far" from x_0. This fact shows that again, the Γ-varying design is an extreme case.

5.5. More technical remarks

• About Assumption K, the first assumption is used to make the kernel K localize the information around the point of estimation x_0 (see (3.2)). The last one is technical and used in the proof of Lemma 6. The two other ones are used for the sake of simplicity, since we only really need the kernel to be bounded from above.

• When $\beta = -1$, Theorem 1 holds only for small regularities $0 < s \le 1$. For technical reasons, we were not able to prove the upper bound when $s > 1$ and $\beta = -1$. More precisely, in this case we have $k = 0$ and in view of (3.4) it is clear that the local polynomial estimator is a Nadaraya–Watson estimator defined by

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

When $s > 1$, we have to use a local polynomial estimator. The problem is then in the asymptotic control of the smallest eigenvalue of $\mathbf{X}_{h_n}^K$ (see Lemma 6) and to do so we use an average (Abelian) transform property of regularly varying functions, which is (see Appendix):

$$\lim_{h\to 0^+} \frac{1}{\ell_\nu(h)} \int y^\alpha K(y)\ell_\nu(yh)\,\frac{dy}{y} = \begin{cases} \int y^{\alpha-1}K(y)\,dy & \text{when } \alpha > 0, \\ +\infty & \text{when } \alpha = 0. \end{cases}$$

Thus the only way to have a limit for both cases is to assume $K(y) = O(|y|^\eta)$ for some $\eta > 0$, but the obtained upper bound rate in this case would be slower than the lower bound.

6. Proofs

6.1. Proof of the main results

Proof of Theorem 1. First we prove the upper bound part of equation (2.4) when $\beta > -1$. We consider the estimator $\widehat{f}_n(x_0) = \widehat{f}_{h_n}(x_0)$, where $\widehat{f}_h$ is given by Definition 4 with h_n given by equation (2.5), and we define $r_n = \omega(h_n)$. Let $0 < \varepsilon \le \frac{1}{2}$. We introduce the event

$$\mathcal{B}_{n,\varepsilon} \triangleq \big\{|\lambda(\mathcal{X}_{h_n}^K) - \lambda_{\beta,K}| \le \varepsilon\big\} \cap \Big\{\Big|\frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1\Big| \le \varepsilon\Big\}.$$

Since $\lim_n nF_\nu(h_n) = +\infty$ (see, for instance, Lemma 4), we have $\mathcal{B}_{n,\varepsilon} \subset \Omega^K_{h_n}$ for n large enough (see (3.11)) and, in particular, on the event $\mathcal{B}_{n,\varepsilon}$ the matrix $\mathbf{X}^K_{h_n}$ is invertible. Then using Proposition 1 and since $f \in \mathcal{F}_{h_n}(x_0, \omega)$, we get:

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

where we last used the definition of h_n. Since, conditionally on $\mathfrak{X}_n$, γ_{h_n} is centered Gaussian such that $\mathbb{E}^n_{f,\mu}\{\gamma^2_{h_n} \mid \mathfrak{X}_n\} \leq 1$, we get for any $p > 0$:

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

where $m(p)$ is defined in Section 4. Now we work on the complement $\mathcal{B}^c_{n,\varepsilon}$. We use Lemmas 2 and 6 to control the probability of $\mathcal{B}_{n,\varepsilon}$ and we recall that $\alpha_n = O(n^\gamma)$ for some $\gamma > 0$. When $N_{n,h_n} = 0$ we have $\widehat{f}_n(x_0) = 0$ by definition and then

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

Then we assume $N_{n,h_n} > 0$. Using Lemma 3 we get:

$$\begin{aligned}&\sup_{f \in \mathcal{U}(\alpha_n)} \mathbb{E}^n_{f,\mu}\big\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p\, \mathbf{1}_{\mathcal{B}^c_{n,\varepsilon}}\big\}\\ &\quad\leq 2^p r_n^{-p}\Big(\sqrt{\mathbb{E}^n_{f,\mu}\{|\widehat{f}_n(x_0)|^{2p}\}} + \alpha_n^p\Big)\sqrt{\mathbb{P}^n_\mu\{\mathcal{B}^c_{n,\varepsilon}\}}\\ &\quad\leq 2^p(\alpha_n r_n^{-1})^p\big(\sqrt{n^p C_{\sigma,k,2p}} + 1\big)\sqrt{\mathbb{P}^n_\mu\{\mathcal{B}^c_{n,\varepsilon}\}} = o_n(1),\end{aligned}$$

and thus we have proved that r_n is an upper bound of the minimax risk (2.4) when $\beta > -1$.

When $\beta = -1$ and $0 < s \leq 1$, we have $k = 0$ and the matrix $\mathcal{X}^K_{h_n}$ is 1×1 sized and equal to $\overline{K}_{n,h_n,0}$ (see equation (6.5)). The bias–variance equation (3.12) becomes in this case:

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

Consider the event

$$\mathcal{C}_{n,\varepsilon} = \Big\{\Big|\frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1\Big| \leq \varepsilon\Big\} \cap \Big\{\Big|\frac{K_{n,h_n,0}}{2nF_\nu(h_n)} - K(0)\Big| \leq \varepsilon\Big\}.$$

We note that the probability of $\mathcal{C}_{n,\varepsilon}$ is controlled by Lemma 2 and equation (6.8) in Lemma 5. Then we can proceed as previously to prove that r_n is an upper bound when $\beta = -1$ and we have proved that r_n is an upper bound for the left-hand side of (2.4). Using Proposition 6 we also have that r_n is a lower bound for the left part of (2.4). The conclusion follows from Lemma 4. □

Proof of Theorem 2. The proof is similar to that of Theorem 1. For the proof of the upper bound part in (2.7) we use the regressogram estimator defined by

$$\widehat{f}_n(x_0) \triangleq \begin{cases} \dfrac{\sum_{i=1}^n Y_i \mathbf{1}_{|X_i - x_0| \leq h_n}}{N_{n,h_n}} & \text{if} \quad N_{n,h_n} > 0, \\ 0 & \text{if} \quad N_{n,h_n} = 0. \end{cases}$$

Let $0 < \varepsilon \leq 1/2$. On the event $\mathcal{D}_{n,\varepsilon} \triangleq \Big\{ \Big| \frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1 \Big| \leq \varepsilon \Big\}$ we clearly have $N_{n,h_n} > 0$ and since $f \in \mathcal{F}_{h_n}(x_0, \omega)$, we have

$$|\widehat{f}_n(x_0) - f(x_0)| \leq \omega(h_n) + \sigma N_{n,h_n}^{-1/2} |v_n| \leq \omega(h_n)(1-\varepsilon)^{-1/2}(1 + |v_n|),$$

where $v_n \triangleq \frac{1}{\sigma \sqrt{N_{n,h_n}}} \sum_{i=1}^n \xi_i \mathbf{1}_{|X_i - x_0| \leq h_n}$ is, conditionally on $\mathfrak{X}_n$, standard Gaussian. Then we get

$$\sup_{f \in \mathcal{F}_{h_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{D}_{n,\varepsilon}} \big\} \leq r_n^p (1-\varepsilon)^{-p/2} m(p).$$

Now we work on $\mathcal{D}_{n,\varepsilon}^c$. If $N_{n,h_n} = 0$, we get using Lemma 2 and since $\alpha_n = O(r_n^{-\gamma})$:

$$\begin{aligned} \sup_{f \in \mathcal{U}(\alpha_n)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_n(x_0) - f(x_0)|^p \, \mathbf{1}_{\mathcal{D}_{n,\varepsilon}^c} \big\} &\leq \alpha_n^p \mathbb{P}_\mu^n \{ \mathcal{D}_{n,\varepsilon}^c \} \\ &= O(r_n^{-\gamma p}) \exp\Big(- \frac{\varepsilon^2 \sigma^2}{1 + \varepsilon/3} r_n^{-2} \Big) = o_n(1), \end{aligned}$$

since $\alpha_n = O(r_n^{-\gamma})$. If $N_{n,h_n} > 0$, since $|\widehat{f}_n(x_0)| \leq \alpha_n + \sigma |v_n|$, we get

$$\sup_{f \in \mathcal{U}(\alpha_n)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{D}_{n,\varepsilon}^c} \big\} \leq 2^p \alpha_n^p (1 + \sqrt{C_{\sigma,0,p}}) \sqrt{\mathbb{P}_\mu^n \{ \mathcal{D}_{n,\varepsilon}^c \}} = o_n(1),$$

where $C_{\sigma,0,p}$ is the same as in the proof of Theorem 1. Thus we have proved that r_n is an upper bound. The lower bound is given by Proposition 6, and the conclusion follows from Lemma 4. □

In the sequel, $\langle \cdot, \cdot \rangle$ denotes the Euclidean scalar product on $\mathbb{R}^{k+1}$, $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{k+1}$, $\| \cdot \|_\infty$ 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 1. On Ω_h^K we have in view of Definition 4 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

$$\sup_{|x - x_0| \leq h} |f(x) - P_f^{n,\varepsilon}(x)| \leq \inf_{P \in \mathcal{P}_k} \sup_{|x - x_0| \leq h} |f(x) - P(x - x_0)| + \frac{\varepsilon}{\sqrt{n}}.$$

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 (3.6)) 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 (3.4) 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 and since $f \in \mathcal{F}_h(x_0, \omega)$, we have:

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

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

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

where we last used the fact that $\|M^{-1}\| = \lambda^{-1}(M)$ for a positive symmetric matrix. The variance term V_h is clearly, conditionally 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, conditionally on $\mathfrak{X}_n$, centered Gaussian of variance:

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

since $K \le 1$. Then $\lambda(\mathcal{X}_h^K) = \inf_{\|x\|=1} \langle x, \mathcal{X}_h^K x\rangle \le \|\mathcal{X}_h^K e_1\| \le \sqrt{k+1}$, since $\mathcal{X}_h^K$ is symmetric and its entries are smaller than 1 in absolute value. Thus

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

and the proposition follows. □

Proof of Proposition 2. The proposition is a direct consequence of Lemmas 1 and 2. □

Proof of Proposition 3. (2) ⇒ (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 $n \to +\infty$ and then $F_\nu \in \mathrm{RV}(\alpha)$ in view of the characterization (A.8) of regular variation. Since $F_\nu(0) = 0$, we have more precisely $F_\nu \in \mathrm{RV}(\alpha)$ for $\alpha \ge 0$ and since ν is monotone, we have $\nu \in \mathrm{RV}(\alpha - 1)$ (see Appendix).

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

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

Then:

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

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

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

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

(1) ⇒ (3): Let $\nu \in \mathrm{RV}(\beta)$ and $0 < \varepsilon \leq 1/2$. If $\beta > -1$, we have $F_\nu \in \mathrm{RV}(\beta+1)$ (see in the Appendix), thus we can write $F_\nu(h) = h^{\beta+1}\ell_F(h)$, where ℓ_F is slowly varying. We define $\gamma_n = n^{-1/(2(\beta+1))}$ when $\beta > -1$ and $\gamma_n = n^{-1}$ if $\beta = -1$. When $\beta = -1$, we have $F_\nu \in \mathrm{RV}(0)$ (see Appendix). We note that in both cases we have $\lim_n \gamma_n = 0$ and $\gamma_{n+1} \sim \gamma_n$ as $n \to +\infty$. In view of Lemma 2 we get for n large enough

$$\mathbb{P}_\mu^n\Big\{ \Big| \frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1 \Big| > \varepsilon \Big\} \leq 2\exp\Big(-\frac{\varepsilon^2}{1+\varepsilon/3}\phi(C)\lambda_n \Big),$$

where we used the fact that ℓ_F 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 $\lim_n n\lambda_n^{-1} = +\infty$ and the proposition follows. □

6.2. Proof of the upper bounds for $\widehat{f}_{H_n}(x_0)$

Proof of Proposition 4. Since $\mathrm{E}_\lambda \subset \Omega_{H_n}^K$, (3.13) and Proposition 1 entail that uniformly in $f \in \mathcal{F}_{H_n}(x_0,\omega)$ we have

$$|\widehat{f}_n(x_0) - f(x_0)| \leq \lambda^{-1}\sqrt{k+1}K_\infty R_n(1 + |\gamma_{H_n}|),$$

where γ_{H_n} is, conditionally on $\mathfrak{X}_n$, centered Gaussian such that $\mathrm{E}_{f,\mu}^n\{\gamma_{H_n}^2 \mid \mathfrak{X}_n\} \leq 1$. The result follows by integration with respect to $\mathbb{P}_{f,\mu}^n(\cdot \mid \mathfrak{X}_n)$. □

Proof of Proposition 5. Let us define $\varepsilon \triangleq \varrho - 1$. We can assume without loss of generality that $\varepsilon < \frac{1}{2} \wedge \lambda_{\beta,K}$. We consider the event $\mathcal{A}_{n,\varepsilon}$ from Lemma 6. In view of this lemma we have $\mathcal{A}_{n,\varepsilon} \subset \mathrm{E}_{\lambda_{\beta,K}-\varepsilon} \cap \{(1-\varepsilon)h_n \leq H_n \leq (1+\varepsilon)h_n\}$ and then $\mathcal{F}_{\varrho h_n}(x_0,\omega) \subset \mathcal{F}_{H_n}(x_0,\omega)$. Thus using Proposition 4 we get

$$\begin{aligned} &\sup_{f \in \mathcal{F}_{\varrho h_n}(x_0,\omega)} \mathrm{E}_{f,\mu}^n\{|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}} \mid \mathfrak{X}_n\} \\ &\quad \leq m(p)(\lambda_{\beta,K} - \varepsilon)^{-p}K_\infty^p(k+1)^{p/2}R_n^p \\ &\quad \leq m(p)(\lambda_{\beta,K} - \varepsilon)^{-p}K_\infty^p(k+1)^{p/2}(1+\varepsilon)^{p(s+1)}r_n^p, \end{aligned}$$

where we used equation (6.1) in the same way as in the proof of Lemma 1 to obtain on $\mathcal{A}_{n,\varepsilon}$ that $\omega(H_n) \leq (1+\varepsilon)^{s+1}\omega(h_n)$. On the complementary event $\mathcal{A}_{n,\varepsilon}^c$, using inequality (6.11) and Lemma 3 and since $\alpha_n = O(n^\gamma)$ for some $\gamma > 0$, we get

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

and (4.2) follows. The equivalent of r_n is given by Lemma 4. $\square$

6.3. Lemmas for the proof of the upper bounds

Lemma 1. *If $\omega \in \mathrm{RV}(s)$ for any $s > 0$, then for any $0 < \varepsilon \leq \frac{1}{2}$ there exists $0 < \eta \leq \varepsilon$ such that*

$$\Big\{ \Big| \frac{N_{n,(1-\varepsilon)h_n}}{2nF_\nu((1-\varepsilon)h_n)} - 1 \Big| \leq \eta \Big\} \cap \Big\{ \Big| \frac{N_{n,(1+\varepsilon)h_n}}{2nF_\nu((1+\varepsilon)h_n)} - 1 \Big| \leq \eta \Big\} \subset \Big\{ \Big| \frac{H_n}{h_n} - 1 \Big| \leq \varepsilon \Big\}.$$

Proof. In view of (3.13) we have $\{H_n \leq (1+\varepsilon)h_n\} = \{N_{n,(1+\varepsilon)h_n} \geq \sigma^2\omega^{-2}((1+\varepsilon)h_n)\}$. Define $\varepsilon_1 \triangleq 1 - (1-\varepsilon^2)^{-2}(1+\varepsilon)^{-2s}$. For ε small enough, it is clear that $\varepsilon_1 > 0$. We recall that ℓ_ω stands for the slowly varying term of ω (see Definition 2). Since (A.1) holds uniformly on each compact set in $(0,+\infty)$, we have for n large enough that for any $y \in [\frac{1}{2}, \frac{3}{2}]$:

$$(1-\varepsilon^2)\ell_\omega(h_n) \leq \ell_\omega(yh_n) \leq (1+\varepsilon^2)\ell_\omega(h_n), \tag{6.1}$$

so using (6.1) with $y = 1+\varepsilon$ $(\varepsilon \leq \frac{1}{2})$, we obtain in view of (2.5):

$$\begin{aligned} 2(1-\varepsilon_1)nF_\nu((1+\varepsilon)h_n) &\geq (1-\varepsilon^2)^{-2}(1+\varepsilon)^{-2s}\sigma^2\omega^{-2}(h_n) \\ &= \sigma^2\big((1+\varepsilon)h_n\big)^{-2s}(1-\varepsilon^2)^{-2}\ell_\omega^{-2}(h_n) \\ &\geq \sigma^2\omega((1+\varepsilon)h_n)^{-2}, \end{aligned}$$

and then

$$\{N_{n,(1+\varepsilon)h_n} \geq 2(1-\varepsilon_1)nF_\nu((1+\varepsilon)h_n)\} \subset \{H_n \leq (1+\varepsilon)h_n\}.$$

Using again (6.1) with $y = 1-\varepsilon$ we get in the same way

$$\{N_{n,(1-\varepsilon)h_n} < 2(1+\varepsilon_1)nF_\nu((1-\varepsilon)h_n)\} \subset \{H_n > (1-\varepsilon)h_n\},$$

and then

$$\Big\{ \Big| \frac{N_{n,(1-\varepsilon)h_n}}{2nF_\nu((1-\varepsilon)h_n)} - 1 \Big| \leq \varepsilon_1 \Big\} \cap \Big\{ \Big| \frac{N_{n,(1+\varepsilon)h_n}}{2nF_\nu((1+\varepsilon)h_n)} - 1 \Big| \leq \varepsilon_1 \Big\} \subset \Big\{ \Big| \frac{H_n}{h_n} - 1 \Big| \leq \varepsilon \Big\}.$$

Now the result follows for the choice $\eta = \varepsilon \wedge \varepsilon_1$. $\square$

Lemma 2. *Under Assumption* M*, we have for any* $\varepsilon, h > 0$*:*

$$\mathbb{P}_\mu^n\Big\{\Big|\frac{N_{n,h}}{2nF_\nu(h)} - 1\Big| > \varepsilon\Big\} \leq 2\exp\Big(-\frac{\varepsilon^2}{1+\varepsilon/3} nF_\nu(h)\Big).$$

Proof. It suffices to apply the Bernstein inequality to the sum of independent random variables $Z_i = \mathbf{1}_{|X_i - x_0| \leq h} - \mathbb{P}_\mu^n\{|X_1 - x_0| \leq h\}$ for $i = 1, \dots, n$. □

Lemma 3. *For any* $p > 0$ *and* $h > 0$ *the estimator* $\widehat{f}_h$ *(see Definition* 4*) satisfies*

$$\sup_{f \in \mathcal{U}(\alpha)} \mathbb{E}_{f,\mu}^n\{|\widehat{f}_h(x_0)|^p \mid \mathfrak{X}_n\} \leq C_{\sigma,k,p}(\alpha\sqrt{n})^p,$$

where $C_{\sigma,k,p} \triangleq (k+1)^{p/2}\sqrt{2/\pi}\int_{\mathbb{R}^+}(1+\sigma t)^p \exp(-t^2/2)\, dt$*.*

Proof. When $N_{n,h} = 0$, we have $\widehat{f}_h = 0$ by definition and the result is obvious, so we assume $N_{n,h} > 0$. Using the fact that $\lambda(A+B) \geq \lambda(A) + \lambda(B)$ when A and B are symmetric non-negative matrices we get $\lambda(\widetilde{\mathbf{X}}_h^K) \geq N_{n,h}^{1/2} > 0$, thus $\widetilde{\mathbf{X}}_h^K$ is invertible. Equation (3.10) 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 $f \in \mathcal{U}(\alpha)$, we have under Assumption K that $|B_{h,j}| \leq \alpha N_{n,h}$, thus $\|B_h\|_\infty \leq \alpha N_{n,h}$. As in the proof of Proposition 1 we have that $\langle(\widetilde{\mathbf{X}}_h^K)^{-1}V_h, e_1\rangle$ is, conditionally on $\mathfrak{X}_n$, centered Gaussian with 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 that all the elements of the matrix $\mathbf{X}_h^K$ are smaller than $N_{n,h}$, thus $\|\mathbf{X}_h^K\| \leq (k+1)N_{n,h}$. Since $\widetilde{\mathbf{X}}_h^K$ is symmetric, we get $\|(\widetilde{\mathbf{X}}_h^K)^{-1}\| = \lambda^{-1}(\widetilde{\mathbf{X}}_h^K) \leq N_{n,h}^{-1/2}$, and then $v_h^2 \leq \sigma^2(k+1)$. Finally, we have

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

where γ_h is, conditionally on $\mathfrak{X}_n$, centered Gaussian with variance smaller than 1. The result follows by integrating with respect to $\mathbb{P}_{f,\mu}^n(\cdot \mid \mathfrak{X}_n)$. □

Lemma 4. *If* $\nu \in \mathrm{RV}(\beta)$*,* $\omega \in \mathrm{RV}(s)$ *for* $s > 0$ *and the sequence* (h_n) *is defined by* (2.5) *then the rate* $r_n = \omega(h_n)$ *satisfies*

$$r_n \sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)}n^{-s/(1+2s+\beta)}\ell_{\omega,\nu}(1/n) \qquad \text{as} \quad n \to +\infty, \tag{6.2}$$

where $\ell_{\omega,\nu}$ *is slowly varying and* $c_{s,\beta} = 4^{s/(1+2s+\beta)}$. *When* $\omega(h) = rh^s$ *(Hölder regularity) for* $r > 0$, *we have more precisely*:

$$r_n \sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{s,\nu}(1/n) \quad \text{as} \quad n \to +\infty, \tag{6.3}$$

where $\ell_{s,\nu}$ *is slowly varying. It is noteworthy that when* $\beta = -1$ *the result becomes*:

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

When $\nu \in \Gamma\mathrm{V}(\rho)$, *we have*

$$r_n \sim \ell_{\omega,\nu}(1/n), \tag{6.4}$$

where $\ell_{\omega,\nu}$ *is slowly varying.*

Proof. Denote $F_\nu(h) \triangleq \int_0^h \nu(t)\,dt$ and let $G(h) = \omega^2(h)F_\nu(h)$. When $\beta > -1$, we have $F_\nu \in \mathrm{RV}(\beta+1)$ (see the Appendix) and when $\beta = -1$, F_ν is slowly varying. Thus $G \in \mathrm{RV}(1+2s+\beta)$ for any $\beta \geq -1$. The function G is continuous and such that $\lim_{h\to 0^+} G(h) = 0$ in view of (A.2), since $1+2s+\beta > 0$. Then, for n large enough, $h_n = G^{\leftarrow}(\sigma^2/(4n))$, where $G^{\leftarrow}(h) \triangleq \inf\{y \geq 0 \mid G(y) \geq h\}$ is the generalized inverse of G. Then in view of (A.8) we have $G^{\leftarrow} \in \mathrm{RV}(1/(1+2s+\beta))$ and then $\omega \circ G^{\leftarrow} \in \mathrm{RV}(s/(1+2s+\beta))$ (see Appendix). Thus we can write $\omega \circ G^{\leftarrow}(h) = h^{s/(1+2s+\beta)}\ell_{\omega,\nu}(h)$, where $\ell_{\omega,\nu}$ is a slowly varying function. Thus:

$$\begin{aligned} r_n = \omega\Big(G^{\leftarrow}\Big(\frac{\sigma^2}{4n}\Big)\Big) &= c_{s,\beta}\sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}\Big(\frac{\sigma^2}{4n}\Big) \\ &\sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(1/n) \qquad \text{as} \quad n \to +\infty, \end{aligned}$$

since ℓ is slowly varying. When $\omega(h) = rh^s$, we can write more precisely $h_n = G^{\leftarrow}(\sigma^2/(4r^2n))$, where $G(h) = h^{2s}F_\nu(h)$, so (6.2) and (6.3) follow.

Let $y \in \mathbb{R}$. Using (A.9) and the uniformity in (A.1) we get $\lim_{h\to 0^+} \ell_\omega(h + y\rho(h))/\ell_\omega(h) = 1$, thus $\lim_{h\to 0^+} \omega(h+y\rho(h))/\omega(h) = 1$. Moreover, since $\Gamma\mathrm{V}(\rho)$ is stable under integration (see Appendix) we have $F_\nu \in \Gamma\mathrm{V}(\rho)$, thus $\lim_{h\to 0^+} G(h + y\rho(y))/G(h) = \exp(y)$ and then $G \in \Gamma\mathrm{V}(\rho)$. For n large enough, h_n is well defined and given by $h_n = G^{\leftarrow}(\sigma^2/(4n))$. Since $G^{\leftarrow} \in \Pi\mathrm{V}(\ell)$ for $\ell = \rho \circ \nu^{\leftarrow} \in \mathrm{RV}(0)$ (see Appendix), $G^{\leftarrow}$ belongs, in particular, to $\mathrm{RV}(0)$ in view of (A.11) and then $r_n = \omega \circ G^{\leftarrow}(\sigma^2/(4n))$, where $\omega \circ G^{\leftarrow} \in \mathrm{RV}(0)$. Thus $r_n \sim \omega \circ G^{\leftarrow}(n^{-1})$ as $n \to +\infty$ and (6.4) follows with $\ell_{\omega,\nu} = \omega \circ G^{\leftarrow}$. □

STUDY OF THE TERMS $\lambda(\mathcal{X}_{h_n}^K)$ AND $\lambda(\mathcal{X}_{H_n}^K)$. We recall that the matrix $\mathcal{X}_{h,K}$ is defined as the symmetric and non-negative matrix with entries $(\mathcal{X}_{h,K})_{j,l} = \overline{K}_{n,h,j+l}$ for $0 \leq j, l \leq k$, where:

$$\overline{K}_{n,h,\alpha} \triangleq \frac{1}{N_{n,h}} \sum_{i=1}^n \Big(\frac{X_i - x_0}{h}\Big)^\alpha K\Big(\frac{X_i - x_0}{h}\Big), \tag{6.5}$$

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

$$K_{\alpha,\beta} \triangleq (1 + (-1)^\alpha) \int_0^1 y^{\alpha+\beta} K(y)\,dy. \tag{6.6}$$

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

$$\mathrm{D}_{n,h,\alpha,K,\varepsilon} \triangleq \Big\{ \Big| \frac{K_{n,h,\alpha}}{nF_\nu(h)} - (\beta+1)K_{\alpha,\beta} \Big| \le \varepsilon \Big\}.$$

Lemma 5. *Let $\alpha \in \mathbb{N}$ and $\varepsilon > 0$. Under Assumption* K *and if $\mu \in \mathcal{R}(x_0,\beta)$ with $\beta > -1$, then for any positive sequence (γ_n) going to 0 we have for n large enough*

$$\mathbb{P}_\mu^n\{\mathrm{D}^c_{n,\gamma_n,\alpha,K,\varepsilon}\} \le 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)} nF_\nu(\gamma_n)\Big). \tag{6.7}$$

When $\beta = -1$ we have:

$$\mathbb{P}_\mu^n\Big\{ \Big| \frac{K_{n,\gamma_n,0}}{nF_\nu(\gamma_n)} - 2K(0) \Big| > \varepsilon \Big\} \le 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)} nF_\nu(\gamma_n)\Big). \tag{6.8}$$

Proof. First we prove (6.7). We define $Q_{i,n,\alpha} \triangleq \big(\frac{X_i - x_0}{\gamma_n}\big)^\alpha K\big(\frac{X_i-x_0}{\gamma_n}\big)$, $Z_{i,n,\alpha} \triangleq Q_{i,n,\alpha} - \mathbb{E}_\mu^n\{Q_{i,n,\alpha}\}$. Since $\mu \in \mathcal{R}(x_0,\beta)$, one has for $i = 1, \dots, n$:

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

where we used Assumption K and the fact that $[x_0 - \gamma_n, x_0 + \gamma_n] \subset W$ for n large enough. Then equations (A.3) and (A.4) entail:

$$\lim_n \frac{1}{nF_\nu(\gamma_n)} \mathbb{E}_\mu^n\{Q_{i,n,\alpha}\} = (\beta+1)K_{\alpha,\beta},$$

and for n large enough:

$$\mathrm{D}^c_{n,\gamma_n,\alpha,K,\varepsilon} \subset \Big\{ \Big| \frac{1}{nF_\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n,\alpha} \Big| > \varepsilon/2 \Big\}. \tag{6.9}$$

In view of Assumption K we have $\mathbb{E}_\mu^n\{Z_{i,n,\alpha}\} = 0$, $|Z_{i,n,\alpha}| \le 2$, and

$$b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}_\mu^n\{Z^2_{i,n,\alpha}\} \le n\mathbb{E}_\mu^n\{Q^2_{1,n,\alpha}\} \le 2nF_\nu(\gamma_n).$$

Since the $Z_{i,n,\alpha}$ are independent, we can apply Bernstein's inequality. If $\tau_n \triangleq \frac{\varepsilon}{2} nF_\nu(\gamma_n)$, equation (6.9) and Bernstein's inequality entail:

$$\mathbb{P}_\mu^n\{\mathrm{D}^c_{n,\gamma_n,\alpha,K,\varepsilon}\} \le 2\exp\Big(\frac{-\tau_n^2}{2(b_n^2 + 2\tau_n/3)}\Big) \le 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)} nF_\nu(\gamma_n)\Big),$$

thus (6.7) follows. The proof of equation (6.8) is similar. When $\beta = -1$, we have $\nu(t) = t^{-1}\ell_\nu(t)$. Define $Z_{i,n} \triangleq Q_{i,n,0} - \mathbb{E}^n_{f,\mu}\{Q_{i,n,0}\}$. In view of equation (A.5) we have

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

Then for n large enough one has

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

The $Z_{i,n}$ are independent and centered and $|Z_{i,n}| \leq 2$. Moreover, in view of Assumption K we have as before $b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}_\mu^n\{Z_{i,n}^2\} \leq 2nF_\nu(\gamma_n)$ and using again the Bernstein inequality we get (6.8). □

Lemma 6. *Let Assumption* K *hold. Assume that* $\omega \in \mathrm{RV}(s)$ *with* $s > 0$, $\mu \in \mathcal{R}(x_0, \beta)$ *with* $\beta > -1$, *and* $\lambda_{\beta,K}$ *is defined by equation* (4.1). *We have* $\lambda_{\beta,K} > 0$ *and for any* $0 < \varepsilon \leq \frac{1}{2}$ *we can find an event* $\mathcal{A}_{n,\varepsilon}$ *such that for* n *large enough*

$$\mathcal{A}_{n,\varepsilon} \subset \{|\lambda(\mathcal{X}_{h_n}^K) - \lambda_{\beta,K}| \leq \varepsilon\} \cap \{|\lambda(\mathcal{X}_{H_n}^K) - \lambda_{\beta,K}| \leq \varepsilon\} \cap \Big\{\Big|\frac{H_n}{h_n} - 1\Big| \leq \varepsilon\Big\} \tag{6.10}$$

and

$$\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\} \leq 4(k+2)\exp\big(-c_{\beta,\sigma,\varepsilon} r_n^{-2}\big), \tag{6.11}$$

where $c_{\beta,\sigma,\varepsilon} > 0$.

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

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

where $x_0 \neq 0$ is the normalized eigenvector associated to the eigenvalue $\lambda_{\beta,K}$ and where we used the fact that

$$\lambda(M) = \inf_{\|x\|=1}\langle x, Mx\rangle, \tag{6.12}$$

for any symmetric matrix M. Then $\forall y \in \operatorname{Supp} K$ 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 $h, \varepsilon > 0$ we introduce the events:

$$\begin{aligned} \mathrm{A}_{n,h,\varepsilon} &= \big\{|\lambda(\mathcal{X}_h^K) - \lambda_{\beta,K}| \leq \varepsilon\big\}, \\ \mathrm{B}_{n,h,\alpha,\varepsilon} &= \Big\{\Big|\overline{K}_{n,h,\alpha} - \frac{\beta+1}{2}K_{\alpha,\beta}\Big| \leq \varepsilon\Big\}. \end{aligned} \tag{6.13}$$

Using the characterization (6.12) we can easily prove that

$$\bigcap_{\alpha=0}^{2k} \mathrm{B}_{n,h,\alpha,\varepsilon/(k+1)^2} \subset \mathrm{A}_{n,h,\varepsilon}. \tag{6.14}$$

Since

$$\overline{K}_{n,H_n,\alpha}-\overline{K}_{n,h_n,\alpha}=\overline{K}_{n,H_n,\alpha}\Big(1-\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^{\alpha}\Big)+\frac{1}{N_{n,h_n}}\sum_{i=1}^{n}\Big(\frac{X_i-x_0}{h_n}\Big)^{\alpha}\Big(K\Big(\frac{X_i-x_0}{H_n}\Big)-K\Big(\frac{X_i-x_0}{h_n}\Big)\Big),$$

we have when K is the rectangular kernel K^R,

$$|\overline{K}_{n,H_n,\alpha}-\overline{K}_{n,h_n,\alpha}|\leq\Big|\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^{\alpha}-1\Big|+\frac{1}{2}\Big(\frac{H_n}{h_n}\vee 1\Big)^{\alpha}\Big|\frac{N_{n,H_n}}{N_{n,h_n}}-1\Big|,$$

and otherwise under Assumption K

$$|\overline{K}_{n,H_n,\alpha}-\overline{K}_{n,h_n,\alpha}|\leq\Big|\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^{\alpha}-1\Big|+\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^{\alpha}\rho\Big|\frac{H_n}{h_n}-1\Big|^{\kappa}+\rho\Big|\frac{h_n}{H_n}-1\Big|^{\kappa}.$$

Let us introduce for $\varepsilon>0$ the event

$$\mathrm{F}_{n,\varepsilon}\triangleq\Big\{\Big|\frac{N_{n,H_n}}{N_{n,h_n}}-1\Big|\leq\varepsilon\Big\}.$$

Then for a good choice of $\varepsilon_1\leq\varepsilon$ we have $|\overline{K}_{n,H_n,\alpha}-\overline{K}_{n,h_n,\alpha}|\leq\frac{\varepsilon}{2(k+1)^2}$ on the event $\mathrm{C}_{n,\varepsilon_1}\cap\mathrm{F}_{n,\varepsilon_1}$ and since $K\leq 1$, we have $K_{\alpha,\beta}\leq\frac{2}{\beta+1}$ and noting that $\mathrm{D}_{n,h,0,K^R,\varepsilon_1}=\big\{\big|\frac{N_{n,h}}{2nF_\nu(h)}-1\big|\leq\varepsilon_1\big\}$, we have for any $\alpha\in\mathbb{N}$

$$\mathrm{D}_{n,h,0,K^R,\frac{\varepsilon}{3(k+1)^2+\varepsilon}}\cap\mathrm{D}_{n,h,\alpha,K,\frac{\varepsilon}{3(k+1)^2+\varepsilon}}\subset\mathrm{B}_{n,h,\alpha,\frac{\varepsilon}{2(k+1)^2}}.$$

Using (6.14) we get for $\eta\triangleq\frac{2\varepsilon}{3(k+1)^2+2\varepsilon}$:

$$\mathrm{D}_{n,h_n,0,K^R,\eta}\cap\bigcap_{\alpha=0}^{2k}\mathrm{D}_{n,h_n,\alpha,K,\eta}\subset\mathrm{A}_{n,h_n,\varepsilon}. \tag{6.15}$$

We take $0<\varepsilon_2\leq\varepsilon_1$ such that $\frac{(1+\varepsilon_2)^{\beta+3}}{1-\varepsilon_2}\leq 1+\varepsilon_1$ (for ε_1 small enough). Since $h\mapsto N_{n,h}$ is increasing we have

$$\mathrm{C}_{n,\varepsilon_2}\subset\{N_{n,(1-\varepsilon_2)h_n}\leq N_{n,H_n}\leq N_{n,(1+\varepsilon_2)h_n}\},$$

and in view of Lemma 1 we can take $0<\varepsilon_3\leq\varepsilon_2$ such that

$$\mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_3}\cap\mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_3}\subset\mathrm{C}_{n,\varepsilon_2}.$$

Using (A.1) with the slowly varying function $\ell_F(h)\triangleq F_\nu(h)h^{-(\beta+1)}$, we have for n large enough that uniformly in $y\in[\frac{1}{2},\frac{3}{2}]$

$$(1-\varepsilon_1)\ell_F(h_n)\leq\ell_F(yh_n)\leq(1+\varepsilon_1)\ell_F(h_n), \tag{6.16}$$

in particular, for $y = 1 - \varepsilon_1$ and $y = 1 + \varepsilon_1$ we get by the definition of ε_2 and since $\varepsilon_3 \le \varepsilon_2 \le \varepsilon_1$:

$$\mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_3} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_3} \cap \mathrm{D}_{n,h_n,0,K^R,\varepsilon_3} \subset \mathrm{F}_{n,\varepsilon_1}.$$

Then we define for $\varepsilon_4 \triangleq \varepsilon_3 \wedge \frac{\varepsilon}{3(k+1)^2+\varepsilon}$ the event

$$\mathcal{A}_{n,\varepsilon} \triangleq \mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,K^R,\varepsilon_4} \cap \bigcap_{\alpha=0}^{2k} \mathrm{D}_{n,h_n,\alpha,K,\varepsilon_4},$$

which satisfies (6.10) in view of the previous embeddings. Using inequality (6.7) in Lemma 5 and since $\varepsilon_4 \le \varepsilon_2 \le \varepsilon_1 \le \frac{1}{2}$, we get

$$\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\} \le 4(k+2)\exp\Big(-\frac{2^{-(\beta+3)}\varepsilon_4\sigma^2}{8(2+\varepsilon_4/3)} r_n^{-2}\Big),$$

where we used (6.16) and (2.5). $\square$

6.4. Proof of the lower bounds

Lemma 7. *If there are two elements f_0 and f_1 of a class Σ 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 $|f_0(x_0) - f_1(x_0)| \ge 2cr_n$ for some constant $c > 0$, then the pointwise minimax risk $\mathcal{R}_n(\Sigma, \mu)$ over the class Σ defined by* (2.1) *in the model* (1.1) *satisfies*:

$$\mathcal{R}_n(\Sigma, \mu) \ge C(c, Q, p) r_n,$$

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

This result is classical. It can be found in Tsybakov (2003) with a proof based on a reduction scheme with two hypotheses and inequalities between the Kullback–Leibler distance and other probability distances.

Proposition 6. *Let h_n be defined by* (2.5)*, let (α_n) be a sequence of positive numbers going to $+\infty$ and $r_n = \omega(h_n)$. If $\Sigma = \Sigma_{h_n,\alpha_n}(x_0, \omega)$ is the class given by Definition* 2*, we have*

$$\liminf_n r_n^{-1} \mathcal{R}_n(\Sigma, \mu) \ge C_{s,p}. \tag{6.17}$$

Proof. We use Lemma 7. 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 $\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n) \le Q$;
(2) $f_{0,n}, f_{1,n} \in \Sigma_{h_n,\alpha_n}(x_0, \omega)$;
(3) $|f_{0,n}(x_0) - f_{1,n}(x_0)| \ge 2cr_n$ for some constant $c > 0$.

We choose the two following hypotheses:

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

(1) Since the ξ_i are centered Gaussian of variance σ^2 and independent of $\mathfrak{X}_n$, we have:

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

then in view of (2.5)

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

(2) For $h \in [0, h_n]$, taking P as the constant polynomial equal to $\omega(h_n)$, we have that the continuity modulus of $f_{0,n}$ is 0, and taking $P = 0$ we obtain that the continuity modulus of $f_{1,n}$ is bounded by $\omega(h)$. Moreover, for n large enough, we clearly have $f_{0,n}, f_{1,n} \in \mathcal{U}(\alpha_n)$ since $\alpha_n \to +\infty$.

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

6.5. Computations of the examples. For a given design density, we compute the minimax convergence rate r_n by first giving an equivalent as $n \to +\infty$ of the smallest solution h_n of

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

and then an equivalent of $r_n = \omega(h_n)$.

6.5.1. *Regularly varying design example.* In the regularly varying design case we find the equivalent of h_n using the following proposition.

Proposition 7. *Let $\gamma > 0$ and $\alpha \in \mathbb{R}$. If $G(h) = h^\gamma (\log(1/h))^\alpha$, then we have:*

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

Proof. When $\alpha = 0$, the result is obvious, hence assume $\alpha \in \mathbb{R} \setminus \{0\}$. 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

$$t \exp(t) = -\gamma x^{1/\alpha}/\alpha, \tag{6.18}$$

where $t \leq 0$. The equation (6.18) 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 in 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 $\lim_{h\to 0^-} W_{-1}(h) = -\infty$. The two solutions of (6.18) are then $t_0 = W_{-1}(-\gamma x^{1/\alpha}/\alpha)$ and $t_1 = W_0(-\gamma x^{1/\alpha}/\alpha)$ and $h_0 \triangleq \exp\big(\alpha W_{-1}(-\gamma x^{1/\alpha}/\alpha)/\gamma\big)$ is the smallest solution. By 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 $x \to 0^+$, 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 $x \to 0^+$.

When $\alpha < 0$, we proceed similarly. We have $t \geq 0$ and (6.18) 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 $x \to +\infty$, we find again $h \sim \gamma^{\alpha/\gamma} x^{1/\alpha}(\log(1/x))^{-\alpha/\gamma}$ as $x \to 0^+$. □

For the second example of regularly varying design, using Proposition 7, we find that an equivalent to the sequence h_n defined by (2.5) is

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

and since $\omega(h) = rh^s(\log(1/h))^\gamma$, we find that an equivalent of r_n (up to a constant depending on s, β, γ, α) is

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

The computation for the third example ($\beta = -1$) is similar to the second example, since $F_\nu(h) = (\log(1/h))^{1-\alpha}$.

6.5.2. *Γ-varying design example.* For the Γ-varying design example $\nu(h) = \exp(-1/h^\alpha)$, we first use the fact that when $\nu \in \Gamma\mathrm{V}(\rho)$, we have $F_\nu(h) \sim \rho(h)\nu(h)$ as $h \to 0^+$ (see Appendix). Recalling that $\rho(h) = \frac{h^{\alpha+1}}{\alpha}$, we solve

$$h^{1+2s+\alpha} \exp(-1/h^\alpha) = y_n, \tag{6.19}$$

where $y_n \triangleq \sigma^2\alpha/(r^2 n)$.

Defining $t \triangleq h^{-\alpha}$, equation (6.19) becomes $t^{-(1+2s+\alpha)/\alpha} \exp(-t) = y_n$, which we rewrite as $x \exp(x) = \alpha/(1 + 2s + \alpha) y_n^{-\alpha/(1+2s+\alpha)}$ for $x \triangleq \alpha/(1 + 2s + \alpha)t$. Then we have $x = W_0\big(\alpha/(1 + 2s + \alpha) y_n^{-\alpha/(1+2s+\alpha)}\big)$, where W_0 is defined in the proof of Proposition 7. Using the fact that $W_0(x) \sim \log(x)$ as $x \to +\infty$, we get $x \sim \frac{\alpha}{1+2s+\alpha} \log n$ as $n \to +\infty$, thus $h_n \sim (\log n)^{-1/\alpha}$ and the result holds since $r_n \triangleq r h_n^s$.

Appendix A. Some Facts on Regular and Γ-Variation

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

A.1. Regular variation. Let ℓ be a slowly varying function throughout the following. An important result is that the property

$$\lim_{h \to 0^+} \ell(yh)/\ell(h) = 1, \tag{A.1}$$

holds *uniformly* for y in any compact set in $(0, +\infty)$. Now if $R_1 \in \mathrm{RV}(\alpha_1)$ and $R_2 \in \mathrm{RV}(\alpha_2)$, one has

(1) $R_1 \times R_2 \in \mathrm{RV}(\alpha_1 + \alpha_2)$,
(2) $R_1 \circ R_2 \in \mathrm{RV}(\alpha_1 \times \alpha_2)$.

If $R \in \mathrm{RV}(\gamma)$ for $\gamma \in \mathbb{R} \setminus \{0\}$, then as $h \to 0^+$ we have

$$R(h) \to \begin{cases} 0 & \text{if} \quad \gamma > 0, \\ +\infty & \text{if} \quad \gamma < 0. \end{cases} \tag{A.2}$$

The asymptotic behaviour of integrals of regularly varying functions, usually called Abelian theorems, plays a key role in the proofs.

• If $\gamma > -1$ we have

$$\int_0^h t^\gamma \ell(t)\, dt \sim (1+\gamma)^{-1} h^{1+\gamma} \ell(h) \qquad \text{as} \quad h \to 0^+, \tag{A.3}$$

and, in particular, $h \mapsto \int_0^h t^\gamma \ell(t)\, dt \in \mathrm{RV}(\gamma+1)$. This result is known as the Karamata theorem.

• When $\gamma = -1$ and if $\int_0^\eta \ell(t) \frac{dt}{t} < +\infty$ for some $\eta > 0$, then $h \mapsto \int_0^h \ell(t) \frac{dt}{t} \in \mathrm{RV}(0)$ and we have

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

• If R is some positive monotone function such that $h \mapsto \int_0^h R(t)\, dt$ belongs to $\mathrm{RV}(\gamma)$ for some $\gamma \geq 0$, then $R \in \mathrm{RV}(\gamma - 1)$.

• If K is a function such that $\int_0^1 t^{-\delta} K(t)\, dt < +\infty$ for some $\delta > 0$, then

$$\int_0^1 K(t) \ell(th)\, dt \sim \ell(h) \int_0^1 K(t)\, dt \qquad \text{as} \quad h \to 0^+. \tag{A.4}$$

Moreover, when $\int_0^\eta \ell(t) dt/t < +\infty$ for some $\eta > 0$, and K is such that $\forall t \geq 0$, $|K(t) - K(0)| \leq \rho |t|^\kappa$ for some $\rho > 0$ and $\kappa > 0$, one has

$$\int_0^1 K(t/h) \ell(t) dt/t \sim K(0) \int_0^1 \ell(t) dt/t \qquad \text{as} \quad h \to 0^+. \tag{A.5}$$

If R is defined and bounded on $[0, +\infty)$, one can define the generalized inverse as

$$R^{\leftarrow}(y) = \inf\{h > 0 \text{ such that } R(h) \geq y\}. \tag{A.6}$$

If $R \in \mathrm{RV}(\gamma)$ for some $\gamma > 0$, then there exists $R^- \in \mathrm{RV}(1/\gamma)$ such that

$$R(R^-(h)) \sim R^-(R(h)) \sim h \qquad \text{as} \quad h \to 0^+, \tag{A.7}$$

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

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 $n \to +\infty$, $\lim_n \delta_n = 0$, and if there is a positive and continuous function ϕ such that for any $y > 0$

$$\lim_n \lambda_n R(y \delta_n) = \phi(y), \tag{A.8}$$

then R varies regularly.

A.2. Γ-VARIATION. We describe now the properties of Γ-varying functions and Π-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 therein.

The first result states that if ν is a function such that (2.6) holds for all $y \in \mathbb{R}$, then (2.6) holds uniformly on each compact set in $\mathbb{R}$. If ρ is such that (2.6) holds, then

$$\lim_{h \to 0^+} \rho(h)/h = 0. \tag{A.9}$$

The auxiliary function ρ in definition (2.6) is unique up to within an asymptotic equivalence and can be taken as $h \mapsto \int_0^h \nu(t)dt/\nu(h)$.

The class $\Gamma\mathrm{V}(\rho)$ is closed under integration. If $\nu \in \Gamma\mathrm{V}(\rho)$, then $F_\nu(h) = \int_0^h \nu(t)\, dt \in \Gamma\mathrm{V}(\rho)$ and we have

$$F_\nu(h) \sim \rho(h)\nu(h) \qquad \text{as} \quad h \to 0^+.$$

We have seen that the class of regularly varying functions RV is closed under the operation of functional inversion. In the case of Γ-variation, the inversion maps the class ΓV in another class of functions, namely the de Haan class ΠV.

Definition 5 (Π-Variation). A function ν is in the de Haan class ΠV if there exists a slowly varying function ℓ and a positive real number c such that

$$\forall y > 0, \quad \lim_{h \to 0^+} (\nu(yh) - \nu(h))/\ell(y) = c \log(y). \tag{A.10}$$

The class of functions ν satisfying (A.10) is denoted by $\Pi\mathrm{V}(\ell)$.

- If $\nu \in \Gamma\mathrm{V}(\rho)$, then $\ell = \rho \circ \nu^{\leftarrow}$ is slowly varying and $\nu^{\leftarrow} \in \Pi\mathrm{V}(\ell)$.
- If $\nu \in \Pi\mathrm{V}(\ell)$ for some $\ell \in \mathrm{RV}(0)$, then $\nu^{\leftarrow} \in \Gamma\mathrm{V}(\rho)$ with $\rho = \ell \circ \nu^{\leftarrow}$.

In both senses the inverses and their auxiliary functions are asymptotically unique. The following inclusion tells us that Π-variation can be viewed as a refinement of slow variation. Actually, any Π-varying function is slowly varying: for any $\ell \in \mathrm{RV}(0)$ we have

$$\Pi\mathrm{V}(\ell) \subset \mathrm{RV}(0). \tag{A.11}$$

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

References

[1] N. H. Bingham, C. M. Goldie, and T. J. Teugels (1989), *Regular Variation*, Encyclopedia of Mathematics and its Applications, Cambridge University Press.

[2] W. S. Cleveland (1979), *Robust locally weighted regression and smoothing scatterplots*, J. Amer. Statist. Soc., **74**, 829–836.

[3] R. M. Corless, G. H. Gonnet, D. E. G. Hare, and D. J. Jeffrey (1996), *On the Lambert W Function*, Adv. Comput. Math., **5**, 329–359.

[4] L. de Haan (1970), *On regular variation and its application to the weak convergence of sample extremes*, PhD Thesis, University of Amsterdam, Mathematical Centre Tract **32**.

[5] J. Fan and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*, Monographs on Statistics and Applied Probability, Chapman & Hall, London.

[6] J. L. Geluk and L. de Haan (1987), *Regular Variations, Extensions and Tauberian Theorems*, CWI Tract.

[7] A. Goldenshluger and A. Nemirovski (1997), *On spatially adaptive estimation of nonparametric regression*, Math. Methods Statist., **6**, 135–170.

[8] E. Guerre (1999), *Efficient random rates for nonparametric regression under arbitrary designs*, Personal communication.

[9] E. Guerre (2000), *Design adaptive nearest neighbor regression estimation*, J. Multivariate Anal., **75**, 219–244.

[10] P. Hall, J. S. Marron, M. H. Neumann, and D. M. Tetterington (1997), *Curve estimation when the design density is low*, Ann. Statist., **25**, 756–770.

[11] I. A. Ibragimov and R. Z. Hasminski (1981), *Statistical Estimation: Asymptotic Theory*, Springer, New York.

[12] J. Karamata (1930), *Sur une mode de croissance régulière des fonctions*, Mathematica (Cluj), **4** 38–53.

[13] V. Korostelev and A. B. Tsybakov (1993), *Minimax Theory of Image Reconstruction*, Springer, New York.

[14] A. Nemirovski (2000), *Topics in Non-Parametric Statistics*, In: *Ecole d'été de probabilités de Saint-Flour XXVIII – 1998*, Lecture Notes in Mathematics, no. 1738, Springer, New York.

[15] S. I. Resnick (1987), *Extreme Values, Regular Variation and Point Processes*, Applied Probability, Springer.

[16] E. Senata (1976), *Regularly Varying Functions*, Lecture Notes in Mathematics, Springer.

[17] V. G. Spokoiny (1998), *Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice*, Ann. Statist., **26**, 1356–1378.

[18] C. J. Stone (1977), *Consistent nonparametric regression*, Ann. Statist., **5**, 595–645.

[19] C. J. Stone (1980), *Optimal rates of convergence for nonparametric estimators*, Ann. Statist., **8**, 1348–1360.

[20] A. B. Tsybakov (1986), *Robust reconstruction of functions by the local approximation*, Problems of Inform. Transmission, **22**, 133–146.

[21] A. B. Tsybakov (2003), *Introduction à l'estimation non-paramétrique*, Springer.

[*Received October* 2004; *revised January* 2005]