

Minimax pointwise estimation of an anisotropic regression function with unknown density of the design

Armelle Guillou & Nicolas Klutchnikoff

CNRS et Université de Strasbourg
Institut de Recherche Mathématique Avancée
UMR 7501
7, rue René Descartes
67084 Strasbourg Cedex, France

July 9, 2010

Abstract. Our aim in this paper is to estimate *as good as possible* an unknown multidimensional regression function at a given point where the design density is also unknown. To reach this goal, we will follow the minimax approach: it will be assumed that the regression function belongs to a known anisotropic Hölder space. On the contrary of the parameters defining the Hölder space, the density of the observations is assumed to be unknown and will be treated as a nuisance parameter. New minimax rates are exhibited as well as local polynomial estimators which achieve these rates. As these estimators depend on a tuning parameter, the problem of its selection is also discussed.

AMS 2000 subject classifications. 62G05, 62G08.

Keywords. Nonparametric regression, anisotropic Hölder spaces, minimax approach, random design, degenerate design.

1 Introduction

In this paper we consider the multidimensional regression problem with a random design. A statistician observes n independent and identically distributed (i.i.d.) observations $\{(X^{(k)}, Y_k)\}_{k=1, \dots, n}$ from the following model:

$$Y_k = f(X^{(k)}) + \sigma \xi_k,$$

where $\{X^{(k)} = (X_1^{(k)}, \dots, X_d^{(k)}) : k = 1 \dots, n\}$ are distributed according to a density function g on $(0, 1)^d$; f is an unknown signal from $(0, 1)^d$ to \mathbb{R} ; $\{\xi_k : k = 1 \dots n\}$ are i.i.d. random variables such that the following conditions are fulfilled: conditionally on \mathcal{X}_n (the σ -algebra engendered by $\{X^{(k)} : k = 1 \dots n\}$) they are distributed according to a standard Gaussian distribution. Under these assumptions, it is clear that the law of the observations is completely determined by the two functions f and g . Here and later we will denote this law by $\mathbb{P}_{f,g}^{(n)}$. Note that, in this regression model, the quantity σ/\sqrt{n} can be viewed as the noise level.

The aim of this paper is to recover the regression function f at a given point $t \in (0, 1)^d$ *as good as possible*. Since $t \in (0, 1)^d$ is assumed to be fixed, we will often omit the dependance on this parameter.

To reach this goal, we will use the minimax approach. Stone (1977) and Ibragimov and Hasminski (1981) were the first who obtained the minimax rate over a non-parametric class of Hölderian functions. As an example, over this class of Hölder functions with smoothness s , Stone (1977) showed that the local polynomial estimator converges with the rate $n^{-s/(1+2s)}$ which is optimal in the minimax sense. However these results are established under the assumption that the design density g is non-vanishing and finite at the point t . Roughly speaking, this means that the information is spatially homogeneous.

Our framework in this paper is different: we consider an inhomogeneous information. This problem has already been considered in Hall *et al.* (1997) who showed that under the assumption that the unknown signal has a Hölder type smoothness of order 2 and if $g(x) \sim x^r$ near 0 where $r > 0$, then a local linear procedure converges with the rate $n^{-4/(5+r)}$ when f is estimated at 0. This rate is also proved to be optimal.

Here we intend to estimate the regression function for random designs—including a large variety of degenerate designs—in a multivariate setting. More precisely, our study combines both minimax and adaptive approaches: the regression function belongs to a known anisotropic Hölder space whereas the density design belongs to a collection of anisotropic regularly varying functions. Thus this paper can be viewed as an extension of Gaïffas (2005) which only considered the univariate case. Due to this new setting, the techniques used are different and our results generalize those obtained by Gaïffas.

Note that the problem of the estimation of a multivariate regression function (with an unknown density of the design) has been also considered in Gaïffas and Lecué (2009) but in an L^2 setting. Their techniques are completely different than ours since they used a minimization of the penalized empirical risk.

Without giving any definition at this point of our paper (see the following subsections), let us explain our goal. If \mathcal{F} denotes a functional space and \mathcal{G} a *good* collection of multivariate distributions, then we want to construct an estimator f_n^* depending only in \mathcal{F} (not in g) such that:

$$\forall g \in \mathcal{G}, \quad \sup_{f \in \mathcal{F}} R_n(f_n^*, f, g) \leq C \sup_{f \in \mathcal{F}} R_n(\tilde{f}_n, f, g)$$

for any estimator \tilde{f}_n if n and C are large enough. The function R_n , called the risk, is used to quantify the quality of any estimator.

1.1 A class for the regression function

In this paper, we suppose that f belongs to an anisotropic Hölder space depending mostly on two parameters s and L where $s = (s_1, \dots, s_d)$ is a smoothness parameter of positive real numbers, and where $L > 0$ is a Lipschitz constant. Each s_i describes the smoothness of f in the i -th direction of the ambient space \mathbb{R}^d . First, let us define, for all $\alpha \in \mathbb{N}^d$ the following polynomial:

$$\begin{aligned} \phi_\alpha : \mathbb{R}^d &\rightarrow \mathbb{R} \\ u &\mapsto \frac{u^\alpha}{\alpha!} := \prod_{i=1}^d \frac{u_i^{\alpha_i}}{\alpha_i!}. \end{aligned}$$

Next, let $\lfloor s \rfloor := (\lfloor s_1 \rfloor, \dots, \lfloor s_d \rfloor)$ where $\lfloor s_i \rfloor$ denotes the largest integer smaller than s_i , i.e., $\lfloor s_i \rfloor = \max\{n \in \mathbb{N} : n < s_i\}$. Then we can introduce the following polynomial space $\mathcal{P}(s)$ viewed as an approximation space:

$$\mathcal{P}(s) := \left\{ \sum_{\alpha \ll \lfloor s \rfloor} a_\alpha \phi_\alpha : a_\alpha \in \mathbb{R} \right\},$$

where $\alpha \ll \alpha' \iff \forall i, \alpha_i \leq \alpha'_i$. This space consists of all anisotropic polynomials of degree less or equal to $\lfloor s \rfloor$.

As we are interested in a pointwise estimation, we will only assume that our function is locally smooth. In order to do that, we will define a family of locally anisotropic Hölder spaces $\mathcal{F}_{i,\delta}^Q(s, L)$. First, let us consider for all $x \in (0, 1)^d$ and all $h \in$

$(0, +\infty)^d$ the set $I(x, h)$ defined by:

$$I(x, h) = \prod_{i=1}^d [x_i - h_i, x_i + h_i]$$

and $\mathcal{H}_x = \{h \in (0, +\infty)^d : I(x, h) \subset (0, 1)^d\}$.

Definition 1 (Locally anisotropic Hölder spaces) Set $t \in (0, 1)^d$, $\delta \in \mathcal{H}_t$ and $Q > 0$. We say that $f : (0, 1)^d \rightarrow \mathbb{R}$ belongs to $\mathcal{F}_{t,\delta}^Q(s, L)$ if:

$$\sup_{x \in I(t,\delta)} |f(x)| \leq Q,$$

and if there exists $P \in \mathcal{P}(s)$ (depending on t and δ) such that:

$$\sup_{u \in I(t,\delta)} \frac{|f(t+u) - P(u)|}{\sum_{i=1}^d |u_i|^{s_i}} \leq L.$$

Using this definition, we can now precise how to quantify the quality of an estimator using an appropriate (family of) risk.

1.2 Minimax rate of convergence

Let us recall that our observations consist in n i.i.d. realizations, of a pair of random variables and that the law of these observations, denoted by $\mathbb{P}_{f,g}^{(n)}$, is completely determined by a regression function $f : (0, 1)^d \rightarrow \mathbb{R}$ and a density function $g : (0, 1)^d \rightarrow \mathbb{R}$. Further we will denote by \mathcal{G} a subclass of all such densities (suitable assumptions on this set will be made later).

As our goal is to estimate f at point $t \in (0, 1)^d$ —note that t can be different from any $X^{(k)}$ —we will quantify the quality of an arbitrary estimator, say \tilde{f}_n , by introducing its risk defined, for any $q > 0$, as follows:

$$R_n(\tilde{f}_n, f, g) := R_n^{(q)}(\tilde{f}_n, f, g) := \left(\mathbb{E}_{f,g}^{(n)} |\tilde{f}_n(t) - f(t)|^q \right)^{1/q},$$

where $\mathbb{E}_{f,g}^{(n)}$ is the expectation under the law $\mathbb{P}_{f,g}^{(n)}$. Unfortunately f and g are unknown; thus this risk cannot be computed in practice.

In order to solve this problem we will adopt a particular approach which mixes a minimax setting—it will be assumed that f belongs to a given functional space

$\mathcal{F}_{t,\delta}^Q(s, L)$ —and an adaptive point of view w.r.t. g which will be treated as a nuisance parameter. More precisely, for all $g \in \mathcal{G}$, the following risk is introduced:

$$R_n(\tilde{f}_n, \mathcal{F}_{t,\delta}^Q(s, L), g) := \sup_{f \in \mathcal{F}_{t,\delta}^Q(s, L)} R_n(\tilde{f}_n, f, g).$$

Since we want to find the *best* estimator, we introduce the following theoretical quantity:

$$R_n(\mathcal{F}_{t,\delta}^Q(s, L), g) := \inf_{\tilde{f}_n} R_n(\tilde{f}_n, \mathcal{F}_{t,\delta}^Q(s, L), g) \quad (1) \quad \boxed{\text{eq:theoreticalRisk}}$$

called the minimax rate of convergence on $\mathcal{F}_{t,\delta}^Q(s, L)$ given $g \in \mathcal{G}$. Note that here the infimum is taken over all the estimators.

Using these definitions, our goal is to construct both a particular estimator, f_n^* (which can depend on s or L but not on g) and a family of normalizations, $\{\varphi_n(s, L, g)\}_{g \in \mathcal{G}}$, such that:

$$\forall g \in \mathcal{G}, \quad \limsup_{n \rightarrow +\infty} \sup_{f \in \mathcal{F}_{t,\delta}^Q(s, L)} \left(\mathbb{E}_{f,g}^{(n)}(\varphi_n^{-1}(s, L, g) |f_n^*(t) - f(t)|)^q \right)^{1/q} < +\infty \quad (\text{U.B.}) \quad \boxed{\text{eq:UB}}$$

and

$$\forall g \in \mathcal{G}, \quad \liminf_{n \rightarrow +\infty} \inf_{\tilde{f}_n} \sup_{f \in \mathcal{F}_{t,\delta}^Q(s, L)} \mathbb{E}_{f,g}^{(n)} \left(\left(\varphi_n^{-1}(s, L, g) |\tilde{f}_n(t) - f(t)| \right)^q \right)^{1/q} > 0. \quad (\text{L.B.}) \quad \boxed{\text{eq:LB}}$$

If (U.B.) and (L.B.) are satisfied, it is easy to prove that the rates $\varphi_n(s, L, g)$ and $R_n(\mathcal{F}_{t,\delta}^Q(s, L), g)$ are asymptotically of the same order. Moreover (U.B.) guarantees that f_n^* achieves the rate $\varphi_n(s, L, g)$ and consequently the minimax rate of convergence.

Our next section is devoted to the construction of a particular procedure of estimation based on local polynomial regression. The performance of this procedure is given in Section 4.

2 Procedure of estimation

2.1 Anisotropic local polynomials

Our aim is to construct an estimator which achieves the rate of convergence $\varphi_n(s, L, g)$, but which does not depend on $g \in \mathcal{G}$ in order to be adaptive w.r.t.

this parameter. To this aim, we construct a *good* local polynomial estimator and derive its upper bound.

More precisely, the idea of the construction is the following: at point $t \in (0, 1)^d$, we choose the best (in an appropriate sense) polynomial in $\mathcal{P}(s)$ to interpolate the Y_k 's in a neighborhood of t .

Remark that if we define the kernel

$$K(x_1, \dots, x_d) := \frac{1}{2^d} \prod_{i=1}^d \mathbb{1}_{[-1,1]}(x_i),$$

then, for all $h \in \mathcal{H}_t$,

$$\widehat{P}_h := \arg \min_{P \in \mathcal{P}(s)} \sum_{k=1}^n \left(Y_k - P(X^{(k)} - t) \right)^2 K\left(\frac{X^{(k)} - t}{h}\right), \quad (2) \quad \boxed{\text{eq:Min1}}$$

is, roughly speaking, the best interpolating polynomial in $\mathcal{P}(s)$ of the observations using points in a h -neighborhood of t . Thus $\widehat{P}_h(0)$ is an estimator of $f(t)$ that we have to compute. To this aim, it is convenient to use some matricial notations.

For any matrix A , we will use A' to denote the transposed matrix of A . Now, let us enumerate the set $\{\alpha \in \mathbb{N}^d : \alpha \ll s\} = \{\alpha^{(1)}, \dots, \alpha^{(\nu)}\}$, where

$$\nu = \prod_{i=1}^d ([s_i] + 1) \quad \text{and} \quad \alpha^{(1)} = (0, \dots, 0).$$

Then define the vectors and matrices

$$\mathbb{Y} = (Y_1, \dots, Y_n)' \in \mathbb{R}^{n \times 1},$$

$$\Omega_h = \begin{pmatrix} K\left(\frac{X^{(1)}-t}{h}\right) & & (0) \\ & \ddots & \\ (0) & & K\left(\frac{X^{(n)}-t}{h}\right) \end{pmatrix} \in \mathbb{R}^{n \times n},$$

and

$$\Phi_h = \begin{pmatrix} \phi_{\alpha^{(1)}}\left(\frac{X^{(1)}-t}{h}\right) & \dots & \phi_{\alpha^{(\nu)}}\left(\frac{X^{(1)}-t}{h}\right) \\ \vdots & & \vdots \\ \phi_{\alpha^{(1)}}\left(\frac{X^{(n)}-t}{h}\right) & \dots & \phi_{\alpha^{(\nu)}}\left(\frac{X^{(n)}-t}{h}\right) \end{pmatrix} \in \mathbb{R}^{n \times \nu}.$$

Using these notations we can re-interpret the minimization problem (2) as follows:

$$\widehat{p}_h = \arg \min_{p \in \mathbb{R}^{\nu \times 1}} \left\| \mathbb{Y} - \Phi_h \begin{pmatrix} p_1 \\ \vdots \\ p_\nu \end{pmatrix} \right\|_h,$$

where

$$\|v\|_h = \sum_{k=1}^n v_k^2 \cdot K\left(\frac{X^{(k)} - t}{h}\right).$$

In fact, the ℓ -th coordinate of \hat{p}_h is exactly the corresponding coefficient of \hat{P}_h multiplied by $\phi_{\alpha^{(\ell)}}(h)$.

The solution of this problem is well-known. It is given by

$$\hat{p}_h = (\Phi'_h \Omega_h \Phi_h)^{-1} \Phi'_h \Omega_h \mathbb{Y},$$

provided $\det(\Phi'_h \Omega_h \Phi_h) > 0$. Therefore, using the fact that $\alpha^{(1)} = (0, \dots, 0)$ —and thus $\phi_{\alpha^{(1)}}(h) = 1$ —it is easy to infer that $\hat{P}_h(0) = e'_1 \hat{p}_h$ where $e'_1 = (1, 0, \dots, 0) \in \mathbb{R}^{1 \times \nu}$. This estimator, denoted as $\hat{f}_{n,h}(t)$, is the local polynomial estimator of degree $\lfloor s \rfloor$ based on the bandwidth h (measurable w.r.t. \mathcal{X}_n). The local polynomial estimator has been extensively studied in the literature (see e.g. Stone (1977), Fan and Gijbels (1996), Spokoiny (1998) among others).

2.2 Selection rule for the bandwidth

In order to complete the definition of our estimator, we need both to construct a data-driven bandwidth and to deal with the case where the associated matrix $(\Phi'_h \Omega_h \Phi_h)$ is not invertible. At this point, we have to introduce some new notations. First, let us introduce the function BV defined for all $h \in (0, +\infty)^d$ by

$$BV(h) = L \sum_{i=1}^d h_i^{s_i} + \frac{\sigma}{\sqrt{N_h}},$$

where N_h is the number of points $X^{(k)}$ in $I(t, h)$. This quantity can be viewed as the sum of the conditional bias term (see Lemma 2) and of the conditional standard deviation (see Lemma 3) of $\hat{f}_{n,h}(t)$. Next, let us define \hat{h}_n as a minimizer of $BV(h)$. The idea is really classical: \hat{h}_n balances the bias and the standard deviation, conditionally on the observations. In other words:

$$\hat{h}_n = \arg \min_{h \in (0, +\infty)^d} BV(h).$$

Let us remark that this quantity can be written as follows:

$$\hat{h}_n = \arg \min_{h \in \{X^{(k)} - t\}_{k=1, \dots, n}} BV(h).$$

Thus \hat{h}_n is measurable w.r.t. the observations and does not depend on g (but depends on s and L).

Using these definitions we are now able to define our estimator as follows:

$$f_n^*(t) = \begin{cases} 0 & \text{if } \lambda_{\min}(M_{\hat{h}_n}) \leq 1/\sqrt{n} \\ \hat{f}_{n,\hat{h}_n}(t) & \text{otherwise} \end{cases}$$

where $\lambda_{\min}(M_h)$ denotes the smallest eigenvalue of the symmetric, positive semi-definite matrix:

$$M_h := \frac{1}{N_h} \Phi_h' \Omega_h \Phi_h.$$

2.3 Heuristic of the construction

First, note that the term $\sigma/\sqrt{N_h}$ can be seen as an empirical version of the following quantity:

$$\frac{\sigma}{\sqrt{n \int_{I(t,h)} g(x) dx}}.$$

Lemma 1 precises this assertion. Thus, the term $BV(h)$ can be viewed as an observable approximation of:

$$\varphi(h) = L \sum_{i=1}^d h_i^{s_i} + \frac{\sigma}{\sqrt{n \int_{I(t,h)} g(x) dx}}$$

which depends explicitly on the unknown function g . It is easy to see that φ achieves its minimum on \mathcal{H}_t . The minimization of this quantity will lead us to consider an ideal bandwidth h_n satisfying the following inequality for a positive constant C :

$$\varphi(h_n) \leq C \min_{h \in \mathcal{H}_t} \varphi(h)$$

which can be used in the case where g is known. Our selection rule is based on the idea that \hat{h}_n can be used instead of h_n . Finally, when $\lambda_{\min}(M_{\hat{h}_n})$ is too small, we affect an arbitrary constant to $f_n^*(t)$. Thus, our estimator will be *good* only if this situation arises with small probability.

Definition 2 *Let us denote by $\varphi_n(s, L, g)$ the infimum of φ . Here, contrary to the function φ , the dependence on s , L and g is explicitly written.*

Remark that our procedure is defined independently on any assumption on the set \mathcal{G} . In order to prove interesting results about this estimator, we must make additional assumptions on this class. In the next section, such assumptions are

presented. Theorem 1 and Theorem 2 give two general results under such assumptions. Nevertheless the rates of convergence exhibited are not explicit. In order to calculate these rates, we will focus our attention on two particular classes of density presented just below.

3 Assumptions on the density function

This section is devoted to the presentation of some *ad hoc* assumptions which will be used in the statement of our results. In the sequel we will use the notations $\widetilde{\mathbb{E}}_{f,g}^{(n)}$ and $\widetilde{\mathbb{P}}_{f,g}^{(n)}$ instead of $\mathbb{E}_{f,g}^{(n)}[\cdot|\mathcal{X}_n]$ and $\mathbb{P}_{f,g}^{(n)}[\cdot|\mathcal{X}_n]$.

3.1 General assumptions

A function $g : (0, 1)^d \rightarrow \mathbb{R}$ is supposed to belong to \mathcal{G} if it is a density function (further \mathcal{D} will denote the set of all such densities) and if it satisfies the following conditions:

$\langle \text{assump:1} \rangle$

Assumption 1 *There exist three positive constants λ , $c_{1,1}$ and $c_{1,2}$ such that, for all $0 < \beta \leq \lambda/2$ and all $h \in (0, 1)^d$ small enough and \mathcal{X}_n -measurable, we have:*

$$\widetilde{\mathbb{P}}_{f,g}^{(n)} [\overline{\Omega}_1(\beta, h)] \leq c_{1,1} \exp \left(-c_{1,2} n \int_{I(t,h)} g(x) dx \right),$$

where $\Omega_1(\beta, h)$ is defined as the event $\{\lambda_{\min}(M_h) > \beta\}$.

$\langle \text{assump:2} \rangle$

Assumption 2 *For all $s \in (0, +\infty)^d$ and $L > 0$, there exist a positive index α and a bandwidth $h_n = (h_{n,1}, \dots, h_{n,d}) \in \mathcal{H}_t$ such that, for all $\gamma > 0$:*

$$\forall i = 1, \dots, d, \quad h_{n,i}^{s_i} \asymp \frac{\sigma/\sqrt{n}}{\left(\int_{I(t,\gamma h_n)} g(x) dx \right)^{1/2}}$$

$\langle \text{assump:3} \rangle$

Assumption 3 *There exist $\alpha > 0$ and a regularly varying function (see e.g. Bingham et al., 1987, p. ??) $\tilde{\varphi}$ of positive index α such that:*

$$\varphi(h_n) \asymp \tilde{\varphi}(\sigma/\sqrt{n}).$$

3.2 Two families of densities

As it will be proved later on, our estimation procedure will chose automatically a bandwidth well-adapted to the unknown density function provided that this density belongs to \mathcal{G} .

In this section, we present two such families of densities. Moreover, for each density on these families we are able to compute the minimax rate of convergence $\varphi_{n(s,L,g)}$ given in Theorems 1 and 2.

When we chose these classes we had in mind to generalize to an high-dimensional setting the class of densities proposed by Gaïffas in the one-dimensional case. In particular, our densities can degenerate in a δ -neighborhood of t (for $\delta \in \mathcal{H}_t$). The degeneration in each direction of the ambient space \mathbb{R}^d is characterized by a real number greater than -1 . Roughly speaking, we will made assumptions such that in the i^{th} direction of \mathbb{R}^d , the unknown density g is regularly varying of index $r_i > -1$. We can interpret this particularity as a loss of information in the direction i if $r_i > 0$, whereas if $r_i \in (-1, 0)$ there is an additional amount of information. The case where r_i is equal to 0 corresponds to the classical uniform distribution. In the sequel r will be defined as the vector $(r_1, \dots, r_d) \in (-1, +\infty)^d$.

Firstly, for all $i = 1, \dots, d$, let us define the following functions:

$$\begin{aligned} G_r^i : \mathbb{R} &\rightarrow \mathbb{R} \\ y &\mapsto |y - t_i|^{r_i} \mathbb{1}_{(0,1)}(y). \end{aligned}$$

Secondly, let us define $G_r^{(1)}$ and $G_r^{(2)}$ as follows:

$$G_r^{(1)}(x) = \prod_{i=1}^d G_r^i(x_i), \quad \text{and} \quad G_r^{(2)}(x) = \sum_{i=1}^d G_r^i(x_i). \quad (3) \quad \boxed{\text{eq:Gr}}$$

Thirdly, let us define $\mathcal{G}_1(r, \delta)$ and $\mathcal{G}_2(r, \delta)$ as follows.

Definition 3 *A function $g : (0, 1)^d \rightarrow \mathbb{R}$ belongs to $\mathcal{G}_1(r, \delta)$ if, and only if, $g \in \mathcal{D}$ and there exists a function $\ell : (0, 1)^d \rightarrow \mathbb{R}$ such that*

$$\forall x \in I(t, \delta), \quad g(x) = G_r^{(1)}(x)\ell(x),$$

and

$$\forall x \in (0, 1)^d, \quad \ell(x) = \prod_{i=1}^d \ell_i(|x_i - t_i|),$$

where, for all $i = 1, \dots, d$, ℓ_i is a slowly varying function at 0 i.e. a positive function such that:

$$\forall \lambda > 0, \quad \lim_{h \rightarrow 0_+} \frac{\ell_i(\lambda h)}{\ell_i(h)} = 1.$$

Definition 4 A function $g : (0, 1)^d \rightarrow \mathbb{R}$ belongs to $\mathcal{G}_2(r, \delta)$ if, and only if, $g \in \mathcal{D}$ and there exists a positive constant c such that:

$$\forall x \in I(t, \delta), \quad g(x) = c G_r^{(2)}(x). \quad (4) \quad \boxed{\text{eq:1}}$$

Finally, let us introduce the following notations:

$$\mathcal{G}_i(\delta) = \bigcup_{r \in (-1, +\infty)^d} \mathcal{G}_i(r, \delta), \quad (i = 1, 2) \quad \text{and} \quad \mathcal{G}_0(\delta) = \mathcal{G}_1(\delta) \cup \mathcal{G}_2(\delta)$$

Let us make some remarks:

- Note that, if g belongs to $\mathcal{G}_1(\delta)$, the components of the $X^{(k)}$'s are independent in a neighborhood of t by assumption. This is not the case if g belongs to $\mathcal{G}_2(\delta)$.
- It is possible to define $\mathcal{G}_2(r, \delta)$ in a more general way by introducing slowly varying functions instead of the constant c in Equation (4). However, for simplicity, we only reduce our study to this simple case here.
- One might also think to another interesting class of densities: the one where, for all x near to t , $g(x)$ could be written as $G^{(3)}(\|x - t\|)$ with $G^{(3)}$ a one-dimensional regularly varying function and $\|\cdot\|$, a given norm on \mathbb{R}^d . In fact, this case could be partially treated thanks to the previous ones.

Indeed, let us consider the case where $G^{(3)}(y) = Cy^r$ for $r, C > 0$ and y near to 0. Since all norms are equivalent we obtain:

$$G^{(3)}(\|x - t\|) \asymp \|x - t\|^r \asymp \|x - t\|_r^r = \sum_{i=1}^d |x_i - t_i|^r \asymp G_{(r, \dots, r)}^{(2)}(x).$$

4 Main results

`<sec:main-results>`

Throughout this section $s \in (0, +\infty)^d$, $L > 0$ and $Q > 0$ are fixed. Moreover, let us consider $\delta_n = (\varphi^{1/s_1}(h_n), \dots, \varphi^{1/s_d}(h_n))$ and let us recall that $\varphi_n(s, L, g)$ is defined as the infimum of φ .

⟨thm:1⟩

Theorem 1 For all density $g \in \mathcal{G}$, there exists a positive constant \bar{c} such that, almost surely:

$$\sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(\widetilde{\mathbb{E}}_{f, g}^{(n)} [|f_n^*(t) - f(t)|^q] \right)^{1/q} \leq \bar{c} \varphi_n(s, L, g),$$

and, consequently:

$$\sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(\mathbb{E}_{f, g}^{(n)} [|f_n^*(t) - f(t)|^q] \right)^{1/q} \leq \bar{c} \varphi_n(s, L, g).$$

⟨thm:2⟩

Theorem 2 For all $g \in \mathcal{G}$, there exists a positive constant \underline{c} such that:

$$\inf_{\tilde{f}_n} \sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(\mathbb{E}_{f, g}^{(n)} \left[\left(\left| \tilde{f}_n(t) - f(t) \right| \right)^q \right] \right)^{\frac{1}{q}} > \underline{c} \varphi_n(s, L, g)$$

where the infimum is taken over all the estimators.

Combining Theorems 1 and 2, it is proved that $\varphi_n(s, L, g)$ is the desired minimax rate of convergence. Remark that this quantity is not explicitly given in these theorems. Nevertheless, as $\varphi_n(s, L, g)$ is the solution of a minimization problem, it is more explicit than $R_n(\mathcal{F}_{t, \delta_n}^Q(s, L), g)$ defined in Equation (1). In fact it is possible to compute, in order, this quantity in many situations using the following proposition.

⟨prop:1⟩

Proposition 1 For all $g \in \mathcal{G}$ we have:

$$\varphi(h_n) \asymp \varphi_n(s, L, g).$$

In particular, we can obtain $\varphi_n(s, L, g)$ explicitly for all $g \in \mathcal{G}_0(\delta_n)$. We have to distinguish two cases if g belongs to $\mathcal{G}_1(\delta_n)$ or $\mathcal{G}_2(\delta_n)$.

⟨prop:UB1⟩

Corollary 1 Set $r \in (-1, +\infty)^d$ and $g \in \mathcal{G}_1(r, \delta_n)$. There exists a slowly varying function at 0, namely ℓ_g , such that, for all n large enough:

$$\varphi_n(s, L, g) \asymp \left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2 \langle \frac{s}{r+1} \rangle}{2 \langle \frac{s}{r+1} \rangle + d}} \ell_g \left(\frac{\sigma}{\sqrt{n}} \right)$$

where

$$\left\langle \frac{s}{r+1} \right\rangle = \left(\frac{1}{d} \sum_{i=1}^d \frac{r_i + 1}{s_i} \right)^{-1}.$$

⟨prop:UB2⟩

Corollary 2 Set $g \in \mathcal{G}_2(\delta_n)$. For all n large enough we have:

$$\varphi_n(s, L, g) \asymp \left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2\langle s \rangle}{2\langle s \rangle + d + \lceil \frac{r \rceil \langle s \rangle}},$$

where

$$\left[\frac{r}{s} \right] = \min_{i=1, \dots, d} \left(\frac{r_i}{s_i} \right).$$

Let us comment these results. In particular we want to point out the following items:

- Our procedure of estimation depends on s and L but does not depend in any way on g nor in the class to which g is supposed to belong. Moreover the optimal bandwidth \hat{h}_n is measurable w.r.t. \mathcal{X}_n .
- If $d = 1$, then the two rates of convergence coincide and are of order:

$$\left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2s}{2s+r+1}}.$$

This is not surprising as $\mathcal{G}_2(\delta_n) \subset \mathcal{G}_1(\delta_n)$, a class already studied by Gaïffas (2005) in dimension 1.

- Let us suppose that $r = 0$ (without any other assumption on s). Then, as in the previous case, the two polynomial parts of the rates coincide and are of order:

$$\left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2\langle s \rangle}{2\langle s \rangle + d}}$$

which is the classical rate when the design is uniformly distributed.

In order to illustrate our theorems we simulate 200 observations from the distribution $G_r^{(1)}$ defined in Equation (3). We consider different values of r and s . Figure 1 represents our observations (points) with the bandwidth (rectangle) \hat{h}_n obtained by our procedure. Note that the point t (set to $(0.5, 0.5)$) is at the center of the rectangle. Although our simulation study is limited, we can draw the following provisional conclusions:

- from top to bottom, for a fixed second component of r set to -0.9 , more smallest is the first component of r (here 3, 0 or -0.9), more concentrated will be the rectangle on the y -axes;

- similarly, if we go from left to right, the first parameter of the regularity decreases (4 and 2), which implies a concentration of the bandwidth on the y -axes;
- as expected the bandwidth is well adjusted to the observations, whatever the regularity of the underlying function f .

To conclude, our estimator is adaptive w.r.t. the density of the design but required the knowledge of the regularity parameters s and L . From a practical point of view this is a problem which can be handle by an adaptive approach (see for instance Lepski (1991), Klutchnikoff (2005) and more recently Goldenshluger and Lepski (2008) and Gaïffas (2009)).

In the next section, we prove our two theorems using some technical lemmas given in Section 6.

5 Proof of our Results

sec:proof-theorems)?

In all the proofs, we denote by C a constant whose value changes all the time and for simplicity let us denote $\varphi_n := \varphi_n(s, L, g)$. Also the notation \bar{A} stands for the complementary of A .

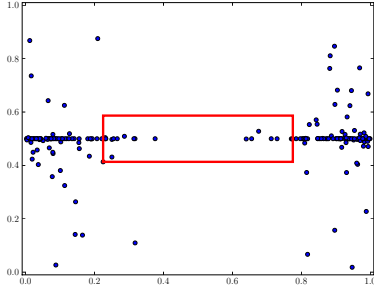
5.1 Proof of Theorem 1

?(sec:proof-thm-1)?

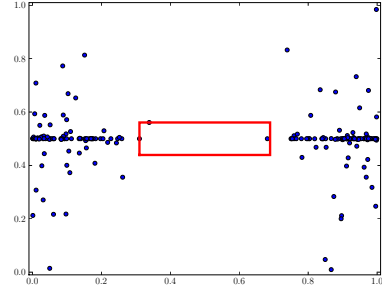
The proof of Theorem 1 is based on some technical lemmas given in the following section. More precisely, Lemmas 1, 2, 3, 4 and 5 are used in this proof. Moreover, Assumptions 1, 2 and 3 are also used in order to prove this result.

Set $s \in (0, +\infty)^d$, $L > 0$, $f \in \mathcal{F}_{t, \delta_n}^Q(s, L)$ and $g \in \mathcal{G}$. Our aim is to control:

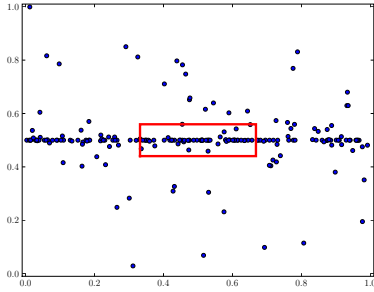
$$(*) = \tilde{\mathbb{E}}_{f, g}^{(n)} [|f_n^*(t) - f(t)|^q].$$



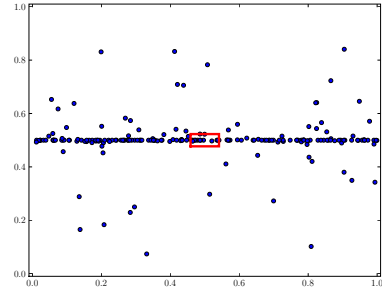
(a)



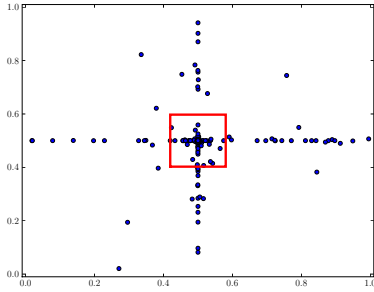
(b)



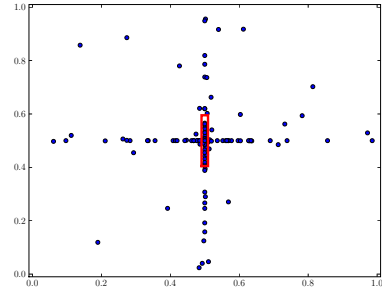
(c)



(d)



(e)



(f)

Figure 1: 200 observations from the distribution $G_r^{(1)}$. The point t is set to $(0.5, 0.5)$. Different values of r and s are chosen: (a) $r = (3, -0.9)$ and $s = (4, 4)$; (b) $r = (3, -0.9)$ and $s = (2, 4)$; (c) $r = (0, -0.9)$ and $s = (4, 4)$; (d) $r = (0, -0.9)$ and $s = (2, 4)$; (e) $r = (-0.9, -0.9)$ and $s = (4, 4)$; (f) $r = (-0.9, -0.9)$ and $s = (2, 4)$.

For n sufficiently large, we obtain:

$$\begin{aligned}
(*) &= \tilde{\mathbb{E}}_{f,g}^{(n)} \left[|f_n^*(t) - f(t)|^q \left(\mathbb{1}_{\overline{\Omega}_1(1/\sqrt{n}, \hat{h}_n)} + \mathbb{1}_{\Omega_1(1/\sqrt{n}, \hat{h}_n)} \right) \right] \\
&= \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n, \hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n}, \hat{h}_n)} \right] + |f(t)|^q \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\overline{\Omega}_1(1/\sqrt{n}, \hat{h}_n) \right] \\
&\leq \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n, \hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n}, \hat{h}_n)} \right] + Q^q \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\overline{\Omega}_1(1/\sqrt{n}, \hat{h}_n) \right] \\
&\leq \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n, \hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n}, \hat{h}_n)} \right] + C \exp \left(-c_{1,2} n \int_{I(t, \hat{h}_n)} g(x) dx \right),
\end{aligned}$$

thanks to Assumption 1. Thus we have now to look at

$$(**) = \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n,\hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n},\hat{h}_n)} \right].$$

Let us bound this term as follows:

$$\begin{aligned} (**) &\leq \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n,\hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(\lambda_0/2,\hat{h}_n) \cap \Omega_2(\eta,h_n)} \right] \\ &\quad + \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n,\hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n},\hat{h}_n) \cap \bar{\Omega}_1(\lambda/2,\hat{h}_n)} \right] \\ &\quad + \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n,\hat{h}_n}(t) - f(t) \right|^q \mathbb{1}_{\Omega_1(1/\sqrt{n},\hat{h}_n) \cap \bar{\Omega}_2(\eta,h_n)} \right] \\ &=: T_{1,n} + T_{2,n} + T_{3,n}. \end{aligned}$$

Control of $T_{1,n}$. Set $0 < \eta < 1$. Under $\Omega_2(\eta, h_n)$ we have:

$$\hat{h}_{n,i}^{s_i} \leq BV(\hat{h}_n) \leq BV(h_n) \leq \varphi(h_n).$$

Thus, written in another way, under this event we have $\hat{h}_n \in I(t, \delta_n)$.

Using the functions $B_q(\cdot)$ and $S_q(\cdot)$ defined in Lemmas 2 and 3, it is clear that

$$\begin{aligned} T_{1,n} &\leq \tilde{\mathbb{E}}_{f,g}^{(n)} \left[(2^{q-1} \vee 1) \left(B_q(\hat{h}_n) + S_q(\hat{h}_n) \right) \mathbb{1}_{\Omega_1(\lambda/2,\hat{h}_n)} \right] \\ &\leq C \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left(L \sum_{i=1}^d \hat{h}_{n,i}^{s_i} \right)^q + \left(\frac{\sigma}{\sqrt{N_{\hat{h}_n}}} \right)^q \right] \\ &\leq C \tilde{\mathbb{E}}_{f,g}^{(n)} \left[BV^q(\hat{h}_n) \right] \\ &\leq C \varphi^q(h_n). \end{aligned}$$

Last inequality holds thanks to Lemma 4.

Control of $T_{2,n}$. From Cauchy-Schwarz's inequality we obtain

$$T_{2,n} \leq \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\sqrt{A_n(\hat{h}_n) C_n(\hat{h}_n)} \right]$$

where

$$\begin{cases} A_n(\hat{h}_n) &= \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \hat{f}_{n,\hat{h}_n}(t) - f(t) \right|^{2q} \mathbb{1}_{\Omega_1(1/\sqrt{n},\hat{h}_n)} \right] \\ C_n(\hat{h}_n) &= \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\bar{\Omega}_1(\lambda/2, \hat{h}_n) \right] \leq c_{1,1} \exp \left(-c_{1,2} n \int_{I(t,\hat{h}_n)} g(x) dx \right). \end{cases}$$

Last upper bound follows from Assumption 1.

Now, let us prove that $A_n(\widehat{h}_n)$ is bounded by n^q (up to a constant). Indeed, using Lemmas 2 and 3, it follows that:

$$\begin{aligned}
A_n(\widehat{h}_n) &\leq C \left\{ \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \widehat{f}_{n,\widehat{h}_n}(t) - \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\widehat{f}_{n,\widehat{h}_n}(t) \right] \right|^{2q} \mathbb{1}_{\Omega_1(1/\sqrt{n},\widehat{h}_n)} \right] \right. \\
&\quad \left. + \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\widehat{f}_{n,\widehat{h}_n}(t) \right] - f(t) \right|^{2q} \mathbb{1}_{\Omega_1(1/\sqrt{n},\widehat{h}_n)} \right] \right\} \\
&\leq C \left\{ n^{q/2} \left(\frac{\sigma}{\sqrt{N_{\widehat{h}_n}}} \right)^{2q} + n^q \left(\sum_{i=1}^d \widehat{h}_{n,i}^{s_i} \right)^{2q} \right\} \\
&\leq C n^q BV^{2q}(\widehat{h}_n) \\
&\leq C n^q BV^{2q}(1, \dots, 1) \leq C n^q.
\end{aligned}$$

Thus:

$$T_{2,n} \leq C n^{q/2} \exp \left(-\frac{c_{1,2}}{2} n \int_{I(t,\widehat{h}_n)} g(x) dx \right)$$

Control of $T_{3,n}$. The computation is very similar to the previous case. But we have to use Lemma 1 instead of Assumption 1 in order to obtain the following bound:

$$T_{3,n} \leq C n^{q/2} \exp \left(-\frac{\Psi_2(\eta)}{2} n \int_{I(t,h_n)} g(x) dx \right).$$

Taking all together, we obtain:

$$\begin{aligned}
(*) &\leq C \left\{ \varphi^q(h_n) + n^{q/2} \left(\exp \left(-\frac{c_{1,2}}{2} n \int_{I(t,\widehat{h}_n)} g(x) dx \right) + \exp \left(-\frac{\Psi_2(\eta)}{2} n \int_{I(t,h_n)} g(x) dx \right) \right) \right\} \\
&\leq C \left\{ \varphi^q(h_n) + n^{q/2} \varphi^\alpha(h_n) \right\}
\end{aligned}$$

for all $\alpha > 0$ thanks to Lemma 5 and a direct computation. Thanks to Assumption 3, there exists $\alpha > 0$ such that, for n large enough, the last term of the right-hand side of this inequality is less than $C\varphi^q(h_n)$. This achieves the proof of our Theorem 1.

5.2 Proof of Theorem 2

(proof-lower-bound)?

Set $s \in (0, +\infty)^d$, $L > 0$ and $g \in \mathcal{G}_1(\delta_n)$. Our goal is to prove the following inequality:

$$\liminf_{n \rightarrow +\infty} \inf_{\tilde{f}_n} \sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(\mathbb{E}_{f, g}^{(n)} \left[\left(\varphi_n^{-1} \left| \tilde{f}_n(t) - f(t) \right| \right)^q \right] \right)^{\frac{1}{q}} > 0$$

where the infimum is taken over all the estimators.

For any $A > 0$, we have:

$$\begin{aligned} & \sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(\mathbb{E}_{f, g}^{(n)} \left[\left(\varphi_n^{-1} \left| \tilde{f}_n(t) - f(t) \right| \right)^q \right] \right)^{\frac{1}{q}} \\ & \geq \sup_{f \in \mathcal{F}_{t, \delta_n}^Q(s, L)} \left(A^q \mathbb{P}_{f, g}^{(n)} \left[\left| \tilde{f}_n(t) - f(t) \right| > A \varphi_n \right] \right)^{\frac{1}{q}} \\ & \geq \max_{f \in \{f_{0, n}, f_{1, n}\}} \left(A^q \mathbb{P}_{f, g}^{(n)} \left[\left| \tilde{f}_n(t) - f(t) \right| > A \varphi_n \right] \right)^{\frac{1}{q}} \\ & \geq A \left(\max_{f \in \{f_{0, n}, f_{1, n}\}} \mathbb{P}_{f, g}^{(n)} \left[\left| \tilde{f}_n(t) - f(t) \right| > A \varphi_n \right] \right)^{\frac{1}{q}}, \end{aligned}$$

where $f_{0, n}$ and $f_{1, n}$ are two functions chosen such that the following assumptions are fulfilled:

1. For $i = 0, 1$, $f_{i, n}$ belongs to $\mathcal{F}_{t, \delta_n}^Q(s, L)$.
2. $|f_{0, n}(t) - f_{1, n}(t)| \geq 2A\varphi_n$.
3. There exists $\alpha > 0$ independent on n such that

$$\mathcal{K} \left(\mathbb{P}_{f_{0, n}, g}^{(n)}, \mathbb{P}_{f_{1, n}, g}^{(n)} \right) \leq \alpha$$

where \mathcal{K} denotes the classical Kullback divergence.

It is well-known, see Theorem 2.2 in Tsybakov (2004, p. 76), that under these conditions we have:

$$\inf_{\tilde{f}_n} \max_{f \in \{f_{0, n}, f_{1, n}\}} \mathbb{P}_{f, g}^{(n)} \left[\left| \tilde{f}_n(t) - f(t) \right| > A \varphi_n \right] \geq \max \left\{ \frac{e^{-\alpha}}{4}, \frac{1 - \sqrt{\alpha/2}}{2} \right\} > 0.$$

Thus, our objective is now to exhibit two such functions (for A properly chosen). In order to achieve this goal, let us consider a function f which belongs to a global

anisotropic Hölder space of parameters s and 1, i.e. a function such that, for all $x \in \mathbb{R}^d$ and all $v \in (0, 1)^d$ there exists a polynomial in $\mathcal{P}(s)$ such that:

$$|f(x + v) - P(v)| \leq \sum_{i=1}^d |v_i|^{s_i}.$$

Moreover, we suppose the following additional conditions on f : $\text{supp } f \subset [-1, 1]^d$ and $\sup_{x \in \mathbb{R}^d} |f(x)| \leq Q$. It is well-known that such a function exists.

Equipped with this definition, let us define

$$f_{0,n} \equiv 0 \quad \text{and} \quad f_{1,n}(x) = L\varphi_n f\left(\dots, \frac{x_i - t_i}{\delta_{n,i}}, \dots\right).$$

Proof of the first assumption. It is clear that $f_{0,n}$ belongs to $\mathcal{F}_{t, \delta_n}^Q(s, L)$. It is also the case for $f_{1,n}$. Then, there exists a polynomial P such that, for all $v \in \mathcal{H}_{\frac{x-t}{\delta_n}}$:

$$\left| f\left(\frac{x-t}{\delta_n} + v\right) - P(v) \right| \leq \sum_{i=1}^d |v_i|^{s_i}.$$

Therefore, by setting $Q(\cdot) := L\varphi_n P\left(\frac{\cdot}{\delta_n}\right)$, we have clearly

$$\begin{aligned} |f_{1,n}(x + u) - Q(u)| &= L\varphi_n \left| f\left(\frac{x-t}{\delta_n} + \frac{u}{\delta_n}\right) - P\left(\frac{u}{\delta_n}\right) \right| \\ &\leq L\varphi_n \sum_{i=1}^d \left| \frac{u_i}{\delta_{n,i}} \right|^{s_i} \\ &\leq L \sum_{i=1}^d |u_i|^{s_i}, \end{aligned}$$

for n large enough thanks to Equations (7).

Proof of the second assumption. By taking $A = L|f(0)|/2$, it is clear that condition $|f_{0,n}(t) - f_{1,n}(t)| \geq 2A\varphi_n$ is fulfilled.

Proof of the third assumption. Let us calculate the Kullback divergence:

$$\begin{aligned} \mathcal{K}(\mathbb{P}_0, \mathbb{P}_1) &= \mathbb{E}_0 \left[\log \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \right] \\ &= \mathbb{E}_0 \left[\mathbb{E}_0 \left[\log \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \middle| \mathcal{X}_n \right] \right], \end{aligned}$$

where \mathbb{E}_0 denotes $\mathbb{E}_{f_{0,n},g}^{(n)}$ and \mathbb{P}_0 and \mathbb{P}_1 are defined similarly. Now, according to Tsybakov (2004, p. 79), we have:

$$\mathbb{E}_0 \left[\log \frac{d\mathbb{P}_0}{d\mathbb{P}_1} \middle| \mathcal{X}_n \right] \leq C \sum_{i=1}^n f_{1,n}^2(X^{(i)}).$$

This implies that:

$$K(\mathbb{P}_0, \mathbb{P}_1) \leq C n \mathbb{E}_0 \left[f_{1,n}^2(X^{(1)}) \right].$$

On the other hand,

$$\begin{aligned} \mathbb{E}_0 \left[f_{1,n}^2(X^{(1)}) \right] &= L^2 \varphi_n^2 \mathbb{E}_0 \left[f^2 \left(\dots, \frac{X_i^{(1)} - t_i}{\delta_{n,i}}, \dots \right) \right] \\ &= L^2 \varphi_n^2 \int_{\mathbb{R}^d} f^2 \left(\frac{x-t}{\delta_n} \right) g(x) dx \\ &\leq L^2 \varphi_n^2 \|f^2\|_\infty \int_{I(t, \delta_n)} g(x) dx \\ &\leq C \frac{\sigma^2 L^2 \|f^2\|_\infty}{n}, \end{aligned}$$

by Assumption 2 and thanks to the fact that f is supported in $[-1, 1]^d$. Thus,

$$K(\mathbb{P}_0, \mathbb{P}_1) \leq C \sigma^2 L^2 \|f^2\|_\infty =: \alpha < +\infty.$$

5.3 Proof of Proposition 1

Firstly, let us notice that there exists $h^* \in (0, +\infty)^d$ such that, for all $h \in (0, +\infty)^d$, $\varphi(h^*) \leq \varphi(h)$. Indeed, it is easy to see that the function φ is continuous on $(0, +\infty)^d$ and satisfies:

$$\lim_{\eta \rightarrow \Delta} \varphi(\eta) = +\infty$$

where $\Delta = \{x \in (0, +\infty)^d : \exists i = 1, \dots, d, \quad x_i = 0\} \cup \{+\infty\}$ is the boundary of $(0, +\infty)^d$. This allows us to conclude.

Secondly, using Assumption 2, we can remark that, for all i :

$$L(h_i^*)^{s_i} \leq L \sum_{j=1}^d (h_j^*)^{s_j} \leq \varphi(h^*) \leq \varphi(h_n) \leq C h_{n,i}^{s_i}.$$

This implies that, for all i :

$$h_i^* \leq \gamma h_{n,i}$$

where $\gamma = \max_{i=1,\dots,d}(C/L)^{1/s_i}$. Thus:

$$\begin{aligned}\varphi(h^*) &\geq \frac{\sigma/\sqrt{n}}{\left(\int_{I(t,h^*)} g(x)dx\right)^{1/2}} \\ &\geq \frac{\sigma/\sqrt{n}}{\left(\int_{I(t,\gamma h_n)} g(x)dx\right)^{1/2}} \\ &\geq C\varphi(h_n).\end{aligned}$$

Last inequality follows again from Assumption 2. On the other hand, thanks to the definition of h^* , the converse inequality ($\varphi(h^*) \leq \varphi(h_n)$) is always satisfied.

Therefore Proposition 1 is proved.

5.4 Proofs of Corollaries 1 and 2

{proof-upper-bound)?

Let us remark that, in order to prove these results, it is sufficient to prove that, if g belongs to $\mathcal{G}_0(\delta_n)$, then Assumptions 1, 2 and 3 are fulfilled. Moreover we have to find h_n as in Proposition 1 and then we have to compute explicitly $\varphi_n(s, L, g)$. These proofs are decomposed into several steps.

Step 1. For all $g \in \mathcal{G}_0(\delta_n)$, Assumption 1 is fulfilled. Indeed, for such a density g , thanks to Lemma 6, there exists M , symmetric and positive definite, satisfying Equation (11). Set $h \in (0, 1)^d$, \mathcal{X}_n -measurable, and $\beta \leq \lambda_{\min}(M)/2$. We have:

$$\tilde{\mathbb{P}}_{f,g}^{(n)} [\bar{\Omega}_0(\beta, h)] = \tilde{\mathbb{P}}_{f,g}^{(n)} [\lambda_{\min}(M_h) \leq \beta] \leq \tilde{\mathbb{P}}_{f,g}^{(n)} [|\lambda_{\min}(M_h) - \lambda_{\min}(M)| > \lambda_{\min}(M)/2].$$

Thanks to Lemma 7, we obtain:

$$\begin{aligned}\tilde{\mathbb{P}}_{f,g}^{(n)} [\bar{\Omega}_0(\beta, h)] &\leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\|M_h - M\|_{\infty} > \frac{\lambda_{\min}(M)}{2C_{2\infty}} \right] \\ &\leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\|M_h - M\|_{\infty} > \frac{\lambda_{\min}(M)}{2 \max(C_{2\infty}, \lambda_{\min}(M))} \right].\end{aligned}$$

And thus:

$$\begin{aligned}
\tilde{\mathbb{P}}_{f,g}^{(n)} [\bar{\Omega}_1(\beta, h)] &\leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\exists (k, \ell) : |(M_h)_{k,\ell} - m_{k,\ell}| > \frac{\lambda_{\min}(M)}{2 \max(C_{2\infty}, \lambda_{\min}(M))} \right] \\
&\leq \sum_{k,\ell=1}^{\nu} \tilde{\mathbb{P}}_{f,g}^{(n)} \left[|(M_h)_{k,\ell} - m_{k,\ell}| > \frac{\lambda_{\min}(M)}{2 \max(C_{2\infty}, \lambda_{\min}(M))} \right] \\
&\leq \nu^2 \tilde{\mathbb{P}}_{f,g}^{(n)} \left[|(M_h)_{k,\ell} - m_{k,\ell}| > \frac{\lambda_{\min}(M)}{2 \max(C_{2\infty}, \lambda_{\min}(M))} \right] \\
&\leq 6\nu^2 \exp \left(-\Psi_1 \left(\frac{\lambda_{\min}(M)}{4 \max(C_{2\infty}, \lambda_{\min}(M))} \right) \left(n \int_{I(t,h)} g(x) dx \right) \right),
\end{aligned}$$

thanks to Equation (11). Taking $c_{1,1} = 6\nu^2$ and $c_{1,2} = \Psi_1 \left(\frac{\lambda_{\min}(M)}{4 \max(C_{2\infty}, \lambda_{\min}(M))} \right)$, we obtain the expected result.

Step 2. Set $r \in (-1, +\infty)^d$. For all $g \in \mathcal{G}_1(r, \delta_n)$, Assumptions 2 and 3 are fulfilled.

Let us consider $h \in \arg \min \varphi$. As φ is differentiable, we have:

$$\forall i = 1, \dots, d, \quad \partial_i \varphi(h) = 0.$$

Let us recall that g can be written in the following form:

$$g(x) = \prod_{i=1}^d g_i(|x_i - t_i|) \quad \text{with} \quad g_i(y) = y^{r_i} \ell_i(y),$$

and ℓ_i a slowly varying function at 0.

Moreover, one can write:

$$\int_{I(t,\eta)} g(x) dx = 2^d \prod_{i=1}^d \int_0^{\eta_i} g_i(x_i) dx_i.$$

Now, let us derivate Equation (14) w.r.t. η_i . Using the previous expression we obtain:

$$\partial_i \varphi(\eta) = L s_i \eta_i^{s_i-1} - \frac{\sigma}{2\sqrt{n}} \frac{1}{\left(\int_{I(t,\eta)} g(x) dx \right)^{1/2}} \frac{g_i(\eta_i)}{\int_0^{\eta_i} g_i(x_i) dx_i}.$$

Moreover

$$\frac{g_i(\eta_i)}{\int_0^{\eta_i} g_i(x_i) dx_i} \tag{5} \boxed{\text{eq:Li}}$$

is regularly varying of index -1 . Thus there exists a slowly varying function at 0, called \mathcal{L}_i such that:

$$(2s_i)^{-1} \frac{g_i(\eta_i)}{\int_0^{\eta_i} g_i(x_i) dx_i} = \frac{1}{\eta_i \mathcal{L}_i(\eta_i)}.$$

Let us remark that Equation (5) combining with Karamata's theorem (see Bingham *et al.*, 1987, p. 26) yields for all $i = 1, \dots, d$:

$$\mathcal{L}_i(\eta_i) \underset{\eta_i \rightarrow 0}{\sim} \frac{1}{1 + r_i}.$$

Now, as $\partial_i \varphi(h) = 0$, we obtain, for all $i = 1, \dots, d$:

$$h_i^{s_i} \mathcal{L}_i(h_i) = \frac{\sigma}{L\sqrt{n}} \frac{1}{\left(\int_{I(t,h)} g(x) dx\right)^{1/2}}. \quad (6) \quad \boxed{\text{eq:toto}}$$

Thus, if $a_n \asymp b_n$ means $0 < \liminf_n \frac{a_n}{b_n} \leq \limsup_n \frac{a_n}{b_n} < +\infty$, we derive:

$$\forall i = 1, \dots, d, \quad h_i^{s_i} \asymp \frac{\sigma}{\sqrt{n}} \left(\int_{I(t,h)} g(x) dx\right)^{-\frac{1}{2}}. \quad (7) \quad \boxed{\text{eq:hni}}$$

At this point, we can see that Assumption 2 is fulfilled. Indeed, thanks to the particular form of g , it is easy to prove that for all $\gamma > 0$,

$$\left(\int_{I(t,h)} g(x) dx\right) \asymp \left(\int_{I(t,\gamma h)} g(x) dx\right).$$

From Equation (6) and using again Karamata's theorem, one can prove that there exists, for all i , a slowly varying function at 0, $\bar{\ell}_i$ depending on g , such that:

$$h_i^{s_i} \mathcal{L}_i(h_i) = \frac{\sigma}{L\sqrt{n}} \frac{1}{\prod_{j=1}^d h_j^{\frac{1+r_j}{2}} \bar{\ell}_j(h_j)}. \quad (8) \quad \boxed{\text{eq:deffhi}}$$

Thus we obtain:

$$\begin{aligned} h_i^{\frac{1+r_i}{2}} \bar{\ell}_i(h_i) &= (h_i^{s_i} \mathcal{L}_i(h_i))^{\frac{1+r_i}{2s_i}} \frac{\bar{\ell}_i(h_i)}{\mathcal{L}_i^{\frac{1+r_i}{2s_i}}(h_i)} \\ &= \left(\frac{\sigma}{L\sqrt{n}} \frac{1}{\prod_{j=1}^d h_j^{\frac{1+r_j}{2}} \bar{\ell}_j(h_j)} \right)^{\frac{1+r_i}{2s_i}} \frac{\bar{\ell}_i(h_i)}{\mathcal{L}_i^{\frac{1+r_i}{2s_i}}(h_i)}. \end{aligned}$$

Taking the product over $i = 1, \dots, d$ in the previous formula, we obtain:

$$\left(\prod_{i=1}^d h_i^{\frac{1+r_i}{2}} \bar{\ell}_i(h_i)\right)^{1+\frac{d}{2\langle\frac{s}{r+1}\rangle}} = \left(\frac{\sigma}{L\sqrt{n}}\right)^{\frac{d}{2\langle\frac{s}{r+1}\rangle}} \prod_{i=1}^d \frac{\bar{\ell}_i(h_i)}{\mathcal{L}_i^{\frac{1+r_i}{2s_i}}(h_i)}.$$

Thus we obtain the following equation:

$$\prod_{i=1}^d h_i^{\frac{1+r_i}{2}} \bar{\ell}_i(h_i) = \left(\frac{\sigma}{L\sqrt{n}} \right)^{\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d}} \left(\prod_{i=1}^d k_i(h_i) \right)^{-1},$$

where

$$k_i(h_i) = \left(\frac{\bar{\ell}_i(h_i)}{\mathcal{L}_i^{\frac{1+r_i}{2s_i}}(h_i)} \right)^{-\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d}}$$

is a slowly varying function. We obtain, by plug-in in Equation (8) the following equality:

$$h_i^{s_i} \mathcal{L}_i(h_i) = \left(\frac{\sigma}{L\sqrt{n}} \right)^{\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d}} \left(\prod_{j=1}^d k_j(h_j) \right). \quad (9) \quad \boxed{\text{eq:touslesmemes}}$$

Let us remark that Equation (9) implies that for all $i \neq j$:

$$h_j^{s_j} \mathcal{L}_j(h_j) = h_i^{s_i} \mathcal{L}_i(h_i) =: \Psi_i(h_i).$$

If we define by Ψ_j^{\leftarrow} the generalized inverse of Ψ_j , then:

$$h_j = \Psi_j^{\leftarrow} \Psi_i(h_i)$$

which is regularly varying of index s_i/s_j . Consequently $\prod_{j=1}^d k_j(h_j) = \tilde{k}_i(h_i)$ which is a slowly varying function at 0. Finally there exists a new slowly varying function $\tilde{\ell}_i$ such that:

$$h_i^{s_i} \tilde{\ell}_i(h_i) = \left(\frac{\sigma}{L\sqrt{n}} \right)^{\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d}}.$$

Thus if we consider the regularly varying function Φ_i of index s_i defined by

$$\Phi_i(x) = x^{s_i} \tilde{\ell}_i(x)$$

it follows that:

$$h_i = \Phi_i^{\leftarrow} \left(\left(\frac{\sigma}{L\sqrt{n}} \right)^{\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d}} \right) =: f_i \left(\frac{\sigma}{L\sqrt{n}} \right)$$

where f_i is a regularly varying function of index $\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d} \frac{1}{s_i}$.

This implies that:

$$h_i = \left(\frac{\sigma}{L\sqrt{n}} \right)^{\frac{2\langle \frac{s}{r+1} \rangle}{2\langle \frac{s}{r+1} \rangle + d} \frac{1}{s_i}} \ell_i^{\#} \left(\frac{\sigma}{\sqrt{n}} \right).$$

As $\varphi_n(s, L, g) \asymp h_i^{s_i}$, we deduce that there exists a regularly varying function $\tilde{\varphi}$ of index $2\langle \frac{s}{r+1} \rangle / (2\langle \frac{s}{r+1} \rangle + d)$ such that:

$$\varphi_n(s, L, g) \asymp \tilde{\varphi} \left(\frac{\sigma}{\sqrt{n}} \right).$$

This proves that Assumption 3 is fulfilled and gives us the order of convergence.

Step 3. Set $r \in (-1, +\infty)^d$ and $g \in \mathcal{G}_2(r, \delta_n)$. Our goal is to find $h_n = (h_{n,1}, \dots, h_{n,d})$ such that Assumption 2 is fulfilled, i.e. such that for all $i \in \{1, \dots, d\}$:

$$h_{n,i}^{s_i} \asymp \frac{\sigma/\sqrt{n}}{\left(\int_{I(t,h)} g(x) dx \right)^{1/2}} \asymp \frac{\sigma/\sqrt{n}}{\left[\left(\prod_{j=1}^d h_j \right) \left(\sum_{k=1}^d h_k^{r_k} \right) \right]^{1/2}}.$$

It is sufficient to consider h_n defined by,

$$\forall i = 1, \dots, d, \quad h_{n,i} = \left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2\langle s \rangle}{2\langle s \rangle + d + \lceil \frac{r \rangle \langle s \rangle} \frac{1}{s_i}}}.$$

Moreover, using this choice it is clear that:

$$\varphi_n(s, L, g) \asymp \left(\frac{\sigma}{\sqrt{n}} \right)^{\frac{2\langle s \rangle}{2\langle s \rangle + d + \lceil \frac{r \rangle \langle s \rangle} \frac{1}{s_i}}}.$$

6 Technical Lemmas

Lemma 1 *If, for all h \mathcal{X}_n -measurable and all $\eta \in (0, 1)$, we define the event:*

$$\Omega_2(\eta, h) := \left\{ \left| \frac{N_h}{n \int_{I(t,h)} g(x) dx} - 1 \right| \leq \eta \right\},$$

then we have, almost surely:

$$1 - \tilde{\mathbb{P}}_{f,g}^{(n)} [\Omega_2(\eta, h)] \leq 2 \exp \left(-\Psi_2(\eta) \left(n \int_{I(t,h)} g(x) dx \right) \right)$$

where $\Psi_2(\eta) := (1 + \eta) (\log(1 + \eta) - 1) + 1$.

Lemma 2 For all h \mathcal{X}_n -measurable and for all $0 < \beta < \lambda/2$, we have, almost surely and uniformly in $f \in \mathcal{F}_{t,\delta_n}^Q(s, L)$ and $g \in \mathcal{G}$:

$$B_q(h) := \left| \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\widehat{f}_{n,h}(t) \right] - f(t) \right|^q \mathbb{1}_{\Omega_1(\beta,h)} \leq C \left(\frac{L}{\beta} \sum_{i=1}^d h_i^{s_i} \right)^q,$$

where $C > 0$.

varianceTermControl)

Lemma 3 For all h \mathcal{X}_n -measurable and for all $0 < \beta < \lambda/2$, we have, almost surely and uniformly in $f \in \mathcal{F}_{t,\delta_n}^Q(s, L)$:

$$S_q(h) := \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \widehat{f}_{n,h}(t) - \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\widehat{f}_{n,h}(t) \right] \right|^q \mathbb{1}_{\Omega_1(\beta,h)} \right] \leq C \left(\frac{\sigma}{\sqrt{\beta N_h}} \right)^q,$$

where $C > 0$ is a suitable constant independent on $g \in \mathcal{G}$.

(lem:psi)

Lemma 4 Set $\psi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ an increasing function such that

$$\psi(x) \geq \exp\left(-\frac{1}{x^2}\right) \quad \forall x > 0. \quad (10) \quad \boxed{\text{condpsi}}$$

Then, there exists a positive constant C such that:

$$\tilde{\mathbb{E}}_{f,g}^{(n)} \left[\psi\left(BV(\widehat{h}_n)\right) \right] \leq C\psi(C\varphi(h_n)).$$

lem:hath-close-to-h)

Lemma 5 For all $\alpha > 0$, for all $g \in \mathcal{G}$, if n is large enough, the following inequality holds:

$$\exp\left(-n \int_{I(t,\widehat{h}_n)} g(x) dx\right) \leq \varphi^\alpha(h_n).$$

o-positive-definite)

Lemma 6 For all $(k, \ell) \in \{1, \dots, \nu\}^2$ and all $\eta \in (0, 1/2)$, if h is \mathcal{X}_n -measurable and converging a.s. to 0, then we have almost surely, for $f \in \mathcal{F}_{t,\delta_n}^Q(s, L)$ and $g \in \mathcal{G}_0(\delta_n)$:

$$\tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| (M_h)_{k,\ell} - m_{k,\ell} \right| \geq \eta \right] \leq 6 \exp\left(-\Psi_1\left(\frac{\eta}{2}\right) \left(n \int_{I(t,h)} g(x) dx\right)\right), \quad (11) \quad \boxed{\text{prelemme}}$$

where $\Psi_1(x) = \frac{x^2}{16 + \frac{4}{3}x}$.

(lem:algebra)

Lemma 7 Let A be a positive semi-definite matrix and B be a positive definite matrix. Then we have

$$|\lambda_{\min}(B) - \lambda_{\min}(A)| \leq C_{2\infty} \|B - A\|_\infty,$$

where $C_{2\infty}$ is such that for all matrix M , $\|M\|_2 \leq C_{2\infty} \|M\|_\infty := \sup_{(i,j)} |M_{i,j}|$.

Proofs of these lemmas are postponed to the appendix.

7 References

?⟨sec:references⟩?

8 Appendix

roofs-techn-lemmas)?

8.1 Proof of Lemma 1

Let us write:

$$\left| \frac{N_h}{n \int_{I(t,h)} g(x) dx} - 1 \right| = \left| \frac{\sum_{k=1}^n \mathbb{1}_{\{X^{(k)} \in I(t,h)\}} - n \int_{I(t,h)} g(x) dx}{n \int_{I(t,h)} g(x) dx} \right|.$$

Moreover, we have:

$$\tilde{\mathbb{E}}_{f,g}^{(n)} \left[\sum_{k=1}^n \mathbb{1}_{\{X^{(k)} \in I(t,h)\}} \right] = n \int_{I(t,h)} g(x) dx = : np_{n,h}.$$

Thus, conditionally on \mathcal{X}_n ,

$$\sum_{k=1}^n \mathbb{1}_{\{X^{(k)} \in I(t,h)\}} \text{ follows a } \mathcal{B}(n, p_{n,h}) \text{ distribution.}$$

It follows that:

$$\tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \frac{N_h}{n \int_{I(t,h)} g(x) dx} - 1 \right| > \eta \right] = \mathbb{P} \left[\frac{|\mathcal{B}(n, p_{n,h}) - np_{n,h}|}{\sqrt{n}} > \eta \sqrt{n} p_{n,h} \right].$$

Applying Bernstein's inequality (see Shorack and Wellner, 1986, p. 440), we obtain:

$$\tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \frac{N_h}{n \int_{I(t,h)} g(x) dx} - 1 \right| > \eta \right] \leq 2 \exp(-np_{n,h} \Psi_2(\eta)),$$

where $\Psi_2(\eta) = (1 + \eta)(\log(1 + \eta) - 1) + 1$. Lemma 1 follows.

8.2 Proof of Lemma 2

As f belongs to the anisotropic Hölder space $\mathcal{F}_{t,\delta_n}^Q(s, L)$, one can write, for all $v \in \mathbb{R}^d$ close to 0 that:

$$f(t + v) - f(t) = P_{f,t}(v) + r_t(v)$$

where $P_{f,t} \in \mathcal{P}(s)$ is such that $P_{f,t}(0) = 0$ and r_t is a remainder term which satisfies:

$$|r_t(u)| \leq L \sum_{i=1}^d |u_i|^{s_i}.$$

Introduce the row vector $Q_h := e'_1 (\Phi'_h \Omega_h \Phi_h)^{-1} \Phi'_h \Omega_h$. Then we have

$$f(t) = Q_h \Phi_h e_1 f(t) = Q_h \begin{pmatrix} f(t) \\ \vdots \\ f(t) \end{pmatrix}.$$

Consequently

$$\widehat{f}_{n,h}(t) - f(t) = Q_h \begin{pmatrix} Y_1 - f(t) \\ \vdots \\ Y_n - f(t) \end{pmatrix} = Q_h \begin{pmatrix} f(X^{(1)}) - f(t) \\ \vdots \\ f(X^{(n)}) - f(t) \end{pmatrix} + \sigma Q_h \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}, \quad (12) \text{ \texttt{decomp}}$$

and we can further decompose the bias term in (12) as

$$Q_h \begin{pmatrix} f(X^{(1)}) - f(t) \\ \vdots \\ f(X^{(n)}) - f(t) \end{pmatrix} = Q_h \begin{pmatrix} P_{f,t}(X^{(1)} - t) \\ \vdots \\ P_{f,t}(X^{(n)} - t) \end{pmatrix} + Q_h \begin{pmatrix} r_t(X^{(1)} - t) \\ \vdots \\ r_t(X^{(n)} - t) \end{pmatrix}.$$

Now, let us remark that there exists $(a_{\alpha(1)}, \dots, a_{\alpha(\nu)}) \in \mathbb{R}^\nu$ such that, for all $v \in \mathbb{R}^d$ we have

$$P_{f,t}(v) = \sum_{\ell=1}^{\nu} a_{\alpha(\ell)} \phi_{\alpha(\ell)}(v).$$

Moreover, $a_{\alpha(1)} = 0$ as $\phi_{\alpha(1)}(\cdot) = 1$ and $\forall \ell \neq 1, \phi_{\alpha(\ell)}(0) = 0$.

Now, our goal is to prove that, for all $\ell \geq 2$

$$Q_h \begin{pmatrix} \phi_{\alpha(\ell)}(X^{(1)} - t) \\ \vdots \\ \phi_{\alpha(\ell)}(X^{(n)} - t) \end{pmatrix} = 0.$$

Indeed, for all $\ell \geq 2$:

$$\phi_{\alpha(\ell)}(h) \cdot \begin{pmatrix} \phi_{\alpha(\ell)}\left(\frac{X^{(1)} - t}{h}\right) \\ \vdots \\ \phi_{\alpha(\ell)}\left(\frac{X^{(n)} - t}{h}\right) \end{pmatrix} = \phi_{\alpha(\ell)}(h) \cdot \Phi_h e_\ell$$

where $e'_\ell = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^{1 \times \nu}$. Thus, thanks to the definition of Q_h we obtain:

$$\begin{aligned} Q_h \begin{pmatrix} \phi_{\alpha^{(\ell)}}(X^{(1)} - t) \\ \vdots \\ \phi_{\alpha^{(\ell)}}(X^{(n)} - t) \end{pmatrix} &= \phi_{\alpha^{(\ell)}}(h) \cdot Q_h \Phi_h e_\ell \\ &= \phi_{\alpha^{(\ell)}}(h) \cdot e'_1 (\Phi'_h \Omega_h \Phi_h)^{-1} (\Phi'_h \Omega_h \Phi_h) e_\ell \\ &= \phi_{\alpha^{(\ell)}}(h) \cdot e'_1 e_\ell = 0. \end{aligned}$$

Finally, using the previous equalities, we deduce that:

$$Q_h \begin{pmatrix} f(X^{(1)}) - f(t) \\ \vdots \\ f(X^{(n)}) - f(t) \end{pmatrix} = Q_h \begin{pmatrix} r_t(X^{(1)} - t) \\ \vdots \\ r_t(X^{(n)} - t) \end{pmatrix} = : \sum_{k=1}^n W_h(X^{(k)} - t) r_t(X^{(k)} - t).$$

Using these notations it is easy to control the bias.

Indeed, if we denote $\mathcal{I}(t, h) = \{k : X^{(k)} \in I(t, h)\}$, we have:

$$\begin{aligned} \left| Q_h \begin{pmatrix} f(X^{(1)}) - f(t) \\ \vdots \\ f(X^{(n)}) - f(t) \end{pmatrix} \right| &\leq \sum_{k=1}^n |W_h(X^{(k)} - t)| \cdot |r_t(X^{(k)} - t)| \\ &\leq \sum_{k \in \mathcal{I}(t, h)} |W_h(X^{(k)} - t)| \cdot |r_t(X^{(k)} - t)| \\ &\leq \sum_{k \in \mathcal{I}(t, h)} |W_h(X^{(k)} - t)| \cdot L \sum_{i=1}^d |X_i^{(k)} - t_i|^{s_i} \\ &\leq \sum_{k \in \mathcal{I}(t, h)} |W_h(X^{(k)} - t)| \cdot L \cdot \sum_{i=1}^d h_i^{s_i} \\ &\leq \left(L \sum_{i=1}^d h_i^{s_i} \right) \sum_{k=1}^n |W_h(X^{(k)} - t)|. \end{aligned}$$

Our aim now is to prove that

$$\sum_{k=1}^n |W_h(X^{(k)} - t)| \leq C. \tag{13} \boxed{\text{borne}}$$

Note that $\sum_{k=1}^n |W_h(X^{(k)} - t)|$ can be bounded by

$$\frac{1}{N_h} \sum_{k=1}^n \left\| e_1' \left(\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h \right)^{-1} \begin{pmatrix} \frac{1}{\alpha^{(1)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(1)}} \\ \vdots \\ \frac{1}{\alpha^{(\nu)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(\nu)}} \end{pmatrix} K \left(\frac{X^{(k)} - t}{h} \right) \right\|_2.$$

By Assumption 1, as we are under $\Omega_1(\beta, h)$, the smallest eigenvalue of $\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h$ is greater than β . Therefore

$$\left\| \left(\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h \right)^{-1} v \right\|_2 \leq \frac{\|v\|_2}{\beta}.$$

Now, the kernel $K(\cdot)$ is defined on $[-1, 1]^d$, null outside, and bounded by 2^{-d} . Therefore

$$\begin{aligned} \sum_{k=1}^n |W_h(X^{(k)} - t)| &\leq \frac{2^{-d}}{\beta N_h} \sum_{k=1}^n \left\| \begin{pmatrix} \frac{1}{\alpha^{(1)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(1)}} \\ \vdots \\ \frac{1}{\alpha^{(\nu)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(\nu)}} \end{pmatrix} \right\|_2 \prod_{i=1}^d \mathbb{1}_{\{t_i - h_i \leq X_i^{(k)} \leq t_i + h_i\}} \\ &\leq \frac{2^{-d}}{\beta N_h} \sum_{k=1}^n \prod_{i=1}^d \mathbb{1}_{\{t_i - h_i \leq X_i^{(k)} \leq t_i + h_i\}} \sqrt{\sum_{\ell=1}^{\nu} \left(\frac{1}{\alpha^{(\ell)}!} \right)^2} \\ &\leq \frac{2^{-d} \sqrt{\nu}}{\beta N_h} \sum_{k \in \mathcal{I}(t, h)} \sqrt{\prod_{i=1}^d \left(\sum_{j=0}^{\lfloor s_i \rfloor} \frac{1}{(j!)^2} \right)} \\ &\leq \frac{\sqrt{\nu}}{\beta}. \end{aligned}$$

Consequently, almost surely, we have the following bound for the bias term:

$$\left| \widetilde{\mathbb{E}}_{f, g}^{(n)} \left[\widehat{f}_{n, h}(t) \right] - f(t) \right| \mathbb{1}_{\Omega_1(\beta, h)} \leq \sqrt{\nu} \left(\frac{L}{\beta} \sum_{i=1}^d h_i^{s_i} \right).$$

Lemma 2 then follows.

8.3 Proof of Lemma 3

Let us write

$$\begin{aligned} Q_h \begin{pmatrix} Y_1 - f(X^{(1)}) \\ \vdots \\ Y_n - f(X^{(n)}) \end{pmatrix} &= \sigma Q_h \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \\ &= \sigma \sum_{k=1}^n W_h(X^{(k)} - t) \xi_k. \end{aligned}$$

Thus, conditionally on \mathcal{X}_n , the variance of this term is equal to:

$$\sigma^2 \sum_{k=1}^n W_h^2(X^{(k)} - t).$$

Now, we would like to bound this variance. To this aim, note that

$$\begin{aligned} \sigma^2 \sum_{k=1}^n W_h^2(X^{(k)} - t) &\leq \sigma^2 \sup_{k=1, \dots, n} |W_h(X^{(k)} - t)| \sum_{k=1}^n |W_h(X^{(k)} - t)| \\ &\leq C \sigma^2 \sup_{k=1, \dots, n} |W_h(X^{(k)} - t)| \end{aligned}$$

by (13). Moreover, under $\Omega_1(\beta, h)$:

$$\begin{aligned} |W_h(X^{(k)} - t)| &\leq \frac{1}{N_h} \left\| e_1' \left(\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h \right)^{-1} \begin{pmatrix} \frac{1}{\alpha^{(1)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(1)}} \\ \vdots \\ \frac{1}{\alpha^{(\nu)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(\nu)}} \end{pmatrix} K \left(\frac{X^{(k)} - t}{h} \right) \right\|_2 \\ &\leq \frac{1}{\beta N_h} \left\| \begin{pmatrix} \frac{1}{\alpha^{(1)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(1)}} \\ \vdots \\ \frac{1}{\alpha^{(\nu)}!} \left(\frac{X^{(k)} - t}{h} \right)^{\alpha^{(\nu)}} \end{pmatrix} K \left(\frac{X^{(k)} - t}{h} \right) \right\|_2 \\ &\leq \frac{1}{\beta N_h}. \end{aligned}$$

This leads to the following result. Conditionally on \mathcal{X}_n we have:

$$\sigma^2 \sum_{k=1}^n W_h^2(X^{(k)} - t) \leq \frac{C \sigma^2}{\beta N_h}.$$

Therefore, Lemma 3 is established for $q = 2$. To conclude with the general case, note that

$$S_q(h) = [S_2(h)]^{q/2} \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left| \frac{\widehat{f}_{n,h}(t) - \tilde{\mathbb{E}}_{f,g}^{(n)}[\widehat{f}_{n,h}(t)]}{\sqrt{S_2(h)}} \right|^q \mathbb{1}_{\Omega_1(\beta,h)} \right] \leq C \left(\frac{\sigma}{\sqrt{\beta N_h}} \right)^q.$$

8.4 Proof of Lemma 4

c:proof-lemma-ref1)?

Set $\eta \in (0, 1)$. As \widehat{h}_n minimizes $BV(\cdot)$ we have:

$$\begin{aligned} \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\psi \left(BV(\widehat{h}_n) \right) \right] &= \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\psi \left(BV(\widehat{h}_n) \right) \left\{ \mathbb{1}_{\Omega_2(\eta, h_n)} + \mathbb{1}_{\overline{\Omega}_2(\eta, h_n)} \right\} \right] \\ &\leq C \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\psi \left(BV(h_n) \right) \mathbb{1}_{\Omega_2(\eta, h_n)} + \psi \left(BV(1, \dots, 1) \right) \mathbb{1}_{\overline{\Omega}_2(\eta, h_n)} \right]. \end{aligned}$$

Now, under $\Omega_2(\eta, h_n)$ we have:

$$BV(h_n) \leq C \left(L \sum_{i=1}^d h_{n,i}^{s_i} + \frac{\sigma}{\sqrt{n \int_{I(t, h_n)} g(x) dx}} \right) \leq C \varphi_n. \quad (14) \text{conv}$$

According to our Lemma 1, this leads to

$$\begin{aligned} \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\psi \left(BV(\widehat{h}_n) \right) \right] &\leq C \left\{ \psi \left(C \varphi_n \right) + \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\overline{\Omega}_2(\eta, h_n) \right] \right\} \\ &\leq C \left\{ \psi \left(C \varphi_n \right) + \exp \left(-n \int_{I(t, h_n)} g(x) dx \right) \right\} \\ &\leq C \left\{ \psi \left(C \varphi_n \right) + \exp \left(-\frac{1}{\varphi_n^2} \right) \right\} \\ &\leq C \psi \left(C \varphi_n \right) \end{aligned}$$

by (10).

8.5 Proof of Lemma 5

We introduce the following notations:

$$\left\{ \begin{array}{l} a_{n,i} = \frac{h_{n,i}}{\tilde{\ell}_i(n)} \quad \text{where } \tilde{\ell}_i(\cdot) \text{ is a suitable slowly varying function tending to infinity} \\ a_n = (a_{n,1}, \dots, a_{n,d}) \\ \Omega_3(i) = \{\widehat{h}_{n,i} \geq a_{n,i}\} \\ \Omega_3 = \bigcap_{i=1}^d \Omega_3(i). \end{array} \right.$$

Therefore

$$\begin{aligned}
\exp\left(-n \int_{I(t, \widehat{h}_n)} g(x) dx\right) &= \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\exp\left(-n \int_{I(t, \widehat{h}_n)} g(x) dx\right) \left\{ \mathbb{1}_{\Omega_3} + \mathbb{1}_{\overline{\Omega}_3} \right\} \right] \\
&\leq \exp\left(-n \int_{I(t, a_n)} g(x) dx\right) + \widetilde{\mathbb{P}}_{f,g}^{(n)}(\overline{\Omega}_3) \\
&\leq \exp\left(-n \int_{I(t, a_n)} g(x) dx\right) + \sum_{i=1}^d \widetilde{\mathbb{P}}_{f,g}^{(n)}(\overline{\Omega}_3(i)) \\
&\leq \exp\left(-n \int_{I(t, a_n)} g(x) dx\right) + \sum_{i=1}^d \widetilde{\mathbb{P}}_{f,g}^{(n)}[\overline{\Omega}_2(\eta, b_n^i)] \\
&\quad + \sum_{i=1}^d \left(\widetilde{\mathbb{P}}_{f,g}^{(n)}[\overline{\Omega}_2(\eta, h_n)] + \widetilde{\mathbb{P}}_{f,g}^{(n)}[\overline{\Omega}_3(i) \cap \Omega_2(\eta, h_n) \cap \Omega_2(\eta, b_n^i)] \right)
\end{aligned}$$

where

$$b_n^i = (b_n^i(1), \dots, b_n^i(d)) \text{ with } b_n^i(j) = \begin{cases} a_{n,i} & \text{if } j = i \\ C \varphi_n^{\frac{1}{s_j}} & \text{otherwise.} \end{cases}$$

All the terms can be handled directly or by using Lemma 1. Moreover, this latter allows us to ensure that under $\Omega_2(\eta, h_n)$ we have

$$L \widehat{h}_{n,i}^{s_i} \leq L \sum_{j=1}^d \widehat{h}_{n,j}^{s_j} \leq C BV(\widehat{h}_n) \leq C BV(h_n) \leq C \varphi_n$$

by (14) and therefore $\widehat{h}_{n,i} \leq C \varphi_n^{\frac{1}{s_i}}$. Thus under $\overline{\Omega}_3(i) \cap \Omega_2(\eta, h_n)$, we have $\widehat{h}_{n,i} \leq C b_n^i$.

Finally, combining all these results, Lemma 5 follows.

8.6 Proof of Lemma 6

This proof will be separated into two cases depending whether g belongs to $\mathcal{G}_1(r, \delta_n)$ or $\mathcal{G}_2(r, \delta_n)$ for a given $r \in (-1, +\infty)^d$.

Case 1. Set $g \in \mathcal{G}_1(r, \delta_n)$. Let us denote by M the matrix defined as follows: for all k, ℓ :

$$m_{k,\ell} = \prod_{i=1}^d \int_{\mathbb{R}} \frac{z_i^{\alpha_i^{(k)} + \alpha_i^{(\ell)}}}{\alpha_i^{(k)}! \alpha_i^{(\ell)}!} \frac{1 + r_i}{4} |z_i|^{r_i} \mathbb{1}_{(-1,1)}(z_i) dz_i.$$

It is easy to prove that M is symmetric (trivial) and positive definite. Indeed, let us denote, for all $z \in \mathbb{R}^d$:

$$\tilde{K}(z) := \prod_{i=1}^d \frac{1+r_i}{4} |z_i|^{r_i} \mathbb{1}_{(-1,1)}(z_i) \geq 0$$

and

$$U(z) := \begin{pmatrix} U_1(z) \\ \vdots \\ U_\nu(z) \end{pmatrix} := \begin{pmatrix} \prod_{i=1}^d \frac{z_i^{\alpha_i^{(1)}}}{\alpha_i^{(1)}!} \\ \vdots \\ \prod_{i=1}^d \frac{z_i^{\alpha_i^{(\nu)}}}{\alpha_i^{(\nu)}!} \end{pmatrix} \in \mathbb{R}^{\nu \times 1}.$$

Using these notations we deduce that:

$$m_{k,\ell} = \int_{\mathbb{R}^d} U_k(z) U_\ell(z) \tilde{K}(z) dz$$

and thus

$$M = (m_{k,\ell})_{k,\ell=1,\dots,\nu} = \int_{\mathbb{R}^d} U(z) U'(z) \tilde{K}(z) dz.$$

In the spirit of Tsybakov (2004, p. 37) we conclude that M is positive definite.

Note, define:

$$\Delta_i(k,\ell) = \phi_{\alpha^{(k)}} \left(\frac{X^{(i)} - t}{h} \right) \phi_{\alpha^{(\ell)}} \left(\frac{X^{(i)} - t}{h} \right) K \left(\frac{X^{(i)} - t}{h} \right).$$

One can re-write $\left(\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h \right)_{k,\ell}$ in the following way:

$$\begin{aligned} \left(\frac{1}{N_h} \Phi_h' \Omega_h \Phi_h \right)_{k,\ell} &= \frac{1}{N_h} \sum_{i=1}^n \Delta_i(k,\ell) \\ &=: \frac{np_{n,h}}{N_h} (Q_{1,n} + Q_{2,n}) \end{aligned}$$

where

$$\begin{aligned} p_{n,h} &:= \int_{I(t,h)} g(x) dx \\ Q_{1,n} &:= \frac{1}{np_{n,h}} \sum_{i=1}^n \tilde{\mathbb{E}}_{f,g}^{(n)} [\Delta_i(k,\ell)] \\ Q_{2,n} &:= \frac{1}{np_{n,h}} \sum_{i=1}^n \left\{ \Delta_i(k,\ell) - \tilde{\mathbb{E}}_{f,g}^{(n)} [\Delta_i(k,\ell)] \right\}. \end{aligned}$$

The two terms will be studied separately. First concerning $Q_{1,n}$, we have:

$$\begin{aligned}
Q_{1,n} &= \frac{1}{\int_{I(t,h)} g(x) dx} \widetilde{\mathbb{E}}_{f,g}^{(n)} \left[\phi_{\alpha^{(k)}} \left(\frac{X^{(1)} - t}{h} \right) \phi_{\alpha^{(\ell)}} \left(\frac{X^{(1)} - t}{h} \right) K \left(\frac{X^{(1)} - t}{h} \right) \right] \\
&= \frac{1}{\int_{I(t,h)} g(x) dx} \int \phi_{\alpha^{(k)}} \left(\frac{u-t}{h} \right) \phi_{\alpha^{(\ell)}} \left(\frac{u-t}{h} \right) K \left(\frac{u-t}{h} \right) g(u) du \\
&= \frac{1}{\int_{I(t,h)} g(x) dx} \int_{[-1,1]^d} \phi_{\alpha^{(k)}}(v) \phi_{\alpha^{(\ell)}}(v) K(v) g(t+vh) \left(\prod_{i=1}^d h_i \right) dv.
\end{aligned}$$

As g belongs to $\mathcal{G}_1(r, \delta_n)$ for a suitable r , we can write g in the following way:

$$g(x) = \prod_{i=1}^d |x_i - t_i|^{r_i} \ell_i(|x_i - t_i|)$$

where ℓ_i is a slowly varying function at 0. Thus:

$$Q_{1,n} = \int_{[-1,1]^d} \frac{\phi_{\alpha^{(k)}}(v) \phi_{\alpha^{(\ell)}}(v) K(v)}{\int_{I(t,h)} g(x) dx} \left(\prod_{i=1}^d |v_i h_i|^{r_i} \ell_i(|v_i h_i|) \right) \left(\prod_{i=1}^d h_i \right) dv.$$

Now, let us compute the following term:

$$\begin{aligned}
\int_{I(t,h)} g(x) dx &= \prod_{i=1}^d \int_{t_i-h_i}^{t_i+h_i} |x_i - t_i|^{r_i} \ell_i(|x_i - t_i|) dx_i \\
&= \prod_{i=1}^d 2 \int_0^{h_i} u_i^{r_i} \ell_i(u_i) du_i \\
&\underset{h \xrightarrow{a.s.} 0}{\sim} \prod_{i=1}^d \frac{2}{1+r_i} h_i^{1+r_i} \ell_i(h_i)
\end{aligned}$$

by Karamata's theorem. Similarly, we obtain:

$$\begin{aligned}
&\int_{[-1,1]^d} \phi_{\alpha^{(k)}}(v) \phi_{\alpha^{(\ell)}}(v) K(v) \left(\prod_{i=1}^d |v_i h_i|^{r_i} \ell_i(|v_i h_i|) \right) \left(\prod_{i=1}^d h_i \right) dv \\
&= \prod_{i=1}^d \frac{h_i^{1+r_i}}{2} \int_{-1}^1 \frac{v_i^{\alpha_i^{(k)} + \alpha_i^{(\ell)}}}{\alpha_i^{(k)}! \alpha_i^{(\ell)}!} |v_i|^{r_i} \ell_i(|v_i h_i|) dv_i \\
&= \prod_{i=1}^d \frac{h_i^{1+r_i}}{2} \frac{1 + (-1)^{\alpha_i^{(k)} + \alpha_i^{(\ell)}}}{\alpha_i^{(k)}! \alpha_i^{(\ell)}!} \int_0^1 v_i^{\alpha_i^{(k)} + \alpha_i^{(\ell)} + r_i} \ell_i(v_i h_i) dv_i \\
&\underset{h \xrightarrow{a.s.} 0}{\sim} \prod_{i=1}^d \frac{h_i^{1+r_i}}{2} \frac{1 + (-1)^{\alpha_i^{(k)} + \alpha_i^{(\ell)}}}{\alpha_i^{(k)}! \alpha_i^{(\ell)}!} \frac{\ell_i(h_i)}{1 + \alpha_i^{(k)} + \alpha_i^{(\ell)} + r_i}
\end{aligned}$$

by Karamata's theorem. Finally:

$$Q_{1,n} \underset{h \xrightarrow{a.s.} 0}{\sim} \prod_{i=1}^d \frac{1+r_i}{4} \frac{1+(-1)^{\alpha_i^{(k)}+\alpha_i^{(\ell)}}}{\alpha_i^{(k)}! \alpha_i^{(\ell)}!} \frac{1}{1+\alpha_i^{(k)}+\alpha_i^{(\ell)}+r_i} = m_{k,\ell}. \quad (15) \quad \boxed{\text{Q1}}$$

Now, let us study $Q_{2,n}$. We have:

- $\tilde{\mathbb{E}}_{f,g}^{(n)}[Q_{2,n}] = 0$.
- For all $i = 1, \dots, n$, we have:

$$|\Delta_i(k, \ell)| \leq \frac{1}{2^d} \frac{1}{\alpha^{(k)}!} \frac{1}{\alpha^{(\ell)}!} \leq \frac{1}{2}.$$

Thus:

$$\left| \Delta_i(k, \ell) - \tilde{\mathbb{E}}_{f,g}^{(n)}[\Delta_i(k, \ell)] \right| \leq 1.$$

- The following inequalities hold

$$\begin{aligned} & \sum_{i=1}^n \tilde{\mathbb{E}}_{f,g}^{(n)} \left[\left(\Delta_i(k, \ell) - \tilde{\mathbb{E}}_{f,g}^{(n)}[\Delta_i(k, \ell)] \right)^2 \right] \\ & \leq n \tilde{\mathbb{E}}_{f,g}^{(n)}[\Delta_1^2(k, \ell)] \\ & \leq n \prod_{i=1}^d \frac{h_i^{1+r_i}}{4} \cdot \frac{2}{\left(\alpha_i^{(k)}! \alpha_i^{(\ell)}! \right)^2} \cdot \frac{\ell_i(h_i)(1+o(1))}{1+2\alpha_i^{(k)}+2\alpha_i^{(\ell)}+r_i} \\ & \leq n \prod_{i=1}^d \frac{h_i^{1+r_i}}{2} \cdot \frac{\ell_i(h_i)(1+o(1))}{1+r_i} \\ & \leq \frac{1}{4^d} n p_{n,h} (1+o(1)). \end{aligned}$$

Thus, applying Bernstein's inequality (see Serfling, 1980, p. 95), we obtain:

$$\tilde{\mathbb{P}}_{f,g}^{(n)}[|Q_{2,n}| \geq \eta] \leq 2 \exp \left(-\frac{\eta^2}{4 + \frac{2}{3}\eta} n p_{n,h} \right). \quad (16) \quad \boxed{\text{Q2}}$$

Finally the above results yield

$$\begin{aligned} & \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \left(\frac{1}{N_h} \Phi'_h \Omega_h \Phi_h \right)_{k,\ell} - m_{k,\ell} \right| \geq \eta \right] \\ & = \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \frac{n p_{n,h}}{N_h} (Q_{1,n} + Q_{2,n}) - m_{k,\ell} \right| \geq \eta \right] \\ & \leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[|Q_{1,n} - m_{k,\ell}| + |Q_{2,n}| \geq \frac{\eta}{2} \right] + \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \left(1 - \frac{N_h}{n p_{n,h}} \right) m_{k,\ell} \right| \geq \left(\frac{N_h}{n p_{n,h}} - \frac{1}{2} \right) \eta \right]. \end{aligned}$$

Then, we can deduce that

$$\begin{aligned} & \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \left(\frac{1}{N_h} \Phi'_h \Omega_h \Phi_h \right)_{k,\ell} - m_{k,\ell} \right| \geq \eta \right] \\ & \leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[|Q_{2,n}| \geq \frac{\eta}{2} - |Q_{1,n} - m_{k,\ell}| \right] + \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\bar{\Omega}_2 \left(\frac{\eta}{2}, h \right) \right] \\ & \quad + \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left\{ \left| \left(1 - \frac{N_h}{np_{n,h}} \right) m_{k,\ell} \right| \geq \left(\frac{N_h}{np_{n,h}} - \frac{1}{2} \right) \eta \right\} \cap \Omega_2 \left(\frac{\eta}{2}, h \right) \right]. \end{aligned}$$

Now, for n sufficiently large, using (15) and the facts that $m_{k,\ell} \in (0, \frac{1}{2})$ and $\eta \in (0, \frac{1}{2})$, we have

$$\begin{aligned} & \tilde{\mathbb{P}}_{f,g}^{(n)} \left[\left| \left(\frac{1}{N_h} \Phi'_h \Omega_h \Phi_h \right)_{k,\ell} - m_{k,\ell} \right| \geq \eta \right] \\ & \leq \tilde{\mathbb{P}}_{f,g}^{(n)} \left[|Q_{2,n}| \geq \frac{\eta}{4} \right] + 2\tilde{\mathbb{P}}_{f,g}^{(n)} \left[\bar{\Omega}_2 \left(\frac{\eta}{2}, h \right) \right] \\ & \leq 2 \exp \left(-\Psi_1 \left(\frac{\eta}{2} \right) np_{n,h} \right) + 4 \exp \left(-\Psi_2 \left(\frac{\eta}{2} \right) np_{n,h} \right) \quad \text{by our Lemma 1 and (16)} \\ & \leq 6 \exp \left(-\Psi_1 \left(\frac{\eta}{2} \right) np_{n,h} \right). \end{aligned}$$

This concludes our first step.

Case 2. Now, let us assume that g belongs to $\mathcal{G}_2(r, \delta_n)$. The only difference with the previous case is that the positive definite matrix M has to be replaced by the following one (which is also positive definite):

$$m_{k,\ell} = \frac{1}{4^d} \left(\prod_{j=1}^d \frac{1 + (-1)^{\alpha_j^{(k)} + \alpha_j^{(\ell)}}}{\alpha_j^{(k)}! \alpha_j^{(\ell)}!} \right) \left(\prod_{j \neq j_0} \frac{1}{1 + \alpha_j^{(k)} + \alpha_j^{(\ell)}} \right) \frac{1 + r_{j_0}}{1 + r_{j_0} + \alpha_{j_0}^{(k)} + \alpha_{j_0}^{(\ell)}},$$

where j_0 is such that $h_{j_0}^{r_{j_0}} = \max_{j=1,\dots,d} h_j^{r_j}$.

8.7 Proof of Lemma 7

Recall that, for all symmetric and positive definite matrix B :

$$\lambda_{\min}(B) = \frac{1}{\rho(B^{-1})}$$

where the spectral radius $\rho(B) = \|B\|_2$ is the greatest eigenvalue of B .

Firstly, let us suppose that A is also a positive definite matrix. In this case we

have:

$$\begin{aligned}
|\lambda_{\min}(B) - \lambda_{\min}(A)| &= \lambda_{\min}(A)\lambda_{\min}(B) |\rho(B^{-1}) - \rho(A^{-1})| \\
&\leq \lambda_{\min}(A)\lambda_{\min}(B) \left| \|B^{-1}\|_2 - \|A^{-1}\|_2 \right| \\
&\leq \lambda_{\min}(A)\lambda_{\min}(B) \|B^{-1} - A^{-1}\|_2 \\
&\leq \lambda_{\min}(A)\lambda_{\min}(B) \|B^{-1}\|_2 \|B - A\|_2 \|A^{-1}\|_2 \\
&\leq \lambda_{\min}(A)\lambda_{\min}(B) \rho(B^{-1})\rho(A^{-1}) \|B - A\|_2 \\
&\leq \|B - A\|_2 \\
&\leq C_{2\infty} \|B - A\|_\infty.
\end{aligned}$$

Consequently, Lemma 7 is proved in this special case.

Secondly, to treat the general case, it is sufficient to construct a sequence $(A_k)_k$ of positive definite matrices such that $\lambda_{\min}(A_k) \rightarrow \lambda_{\min}(A) = 0$ and $A_k \rightarrow A$ when $k \rightarrow \infty$ and to use continuity arguments.