Minimax and minimax adaptive estimation in multiplicative regression: locally bayesian approach

M. Chichignoud

Université Aix-Marseille 1
LATP, 39 rue Joliot Curie, 13453 Marseille cedex 13, FRANCE,
e-mail: chichign@cmi.univ-mrs.fr

Abstract: This paper deals with the non-parametric estimation in the regression with the multiplicative noise. Using the local polynomial fitting and the bayesian approach, we construct the minimax on isotropic Hölder class estimator. Next, applying Lepski's method we propose an estimator which is optimally adaptive over the collection of isotropic Hölder classes. To prove the optimality of the proposed procedure, we establish in particular the exponential inequality for the deviation of locally bayesian estimator since the parameter estimated. These theoretical results are illustrated by simulation study.

AMS 2000 subject classifications: 62G08, 62G20.
Keywords and phrases: local bayesian fitting, multiplicative regression, adaptive bandwidth selector, Lepski's method, optimality criterion.

1. Introduction

Let the statistical experiment be generated by the couples of observations $Y^{(n)} = (X_i, Y_i)_{i=1,\dots n}$, $n \in \mathbb{N}^*$ where (X_i, Y_i) satisfies the equation

$$Y_i = f(X_i) \times U_i, \quad i = 1, \dots, n. \tag{1.1}$$

Here $f : [0,1]^d \to \mathbb{R}$ is an unknown function and we are interested in estimating f at a given point $y \in [0,1]^d$ from the observation $Y^{(n)}$.

The random variables (the noise) $(U_i)_{i \in 1,\dots,n}$ are supposed to be independent and uniformly distributed on $[0,1]$.

The design points $(X_i)_{i \in 1,\dots,n}$ are deterministic and without loss of generality we will assume that

$$X_i \in \left\{ 1/n^{1/d}, 2/n^{1/d}, \dots, 1 \right\}^d, \ i = 1, \dots, n.$$

All along the paper the unknown function f is supposed to be smooth, in particular, it belongs to the Hölder ball of functions $\mathbb{H}_d(\beta, L, M)$ (cf. Definition 1 below). Here $\beta > 0$ is the smoothness of f, M is the sum of upper bounds of f and its partial derivatives and $L > 0$ is the Lipschitz constant.

Moreover, we will consider only the functions f separated away from zero by some positive constant. From now on we will suppose that there exists $0 < A < M$ such that $f \in \mathbb{H}_d(\beta, L, M, A)$, where

$$\mathbb{H}_d(\beta, L, M, A) = \left\{ g \in \mathbb{H}_d(\beta, L, M) : \inf_{x \in [0,1]^d} g(x) \geq A \right\}.$$

Motivation. The theoretical interest to the multiplicative regression model (1.1) with discontinuous noise is dictated by the following fact. The typical approach to the study of the models with multiplicative noise consists in their transformation into the model with an additive noise. After that in the application, the linear smoothing technic are used, based on standard methods like kernel smoothing, local polynomials, etc. But these technics are not always optimal. Let us illustrate the latter approach by the consideration of one of the most popular non-parametric model namely multiplicative gaussian regression

$$Y_i = \sigma(X_i)\, \xi_i, \quad i = 1, \ldots, n. \tag{1.2}$$

Here $(\xi_i)_{i=1,\ldots,n}$ are i.i.d. standard gaussian random variables and the goal is to estimate the variance $\sigma^2(\cdot)$.

Set $Y_i' = Y_i^2$ and $\eta_i = \xi_i^2 - 1$, one can transform the model (1.2) into the heteroscedastic additive regression:

$$Y_i' = \sigma^2(X_i) + \sigma^2(X_i)\, \eta_i, \quad i = 1, \ldots, n,$$

where, obviously, $\mathbb{E}\eta_i = 0$. Applying any of the linear methods mentioned above to the estimation of $\sigma^2(\cdot)$, one can construct an estimator whose estimation accuracy is given by $n^{-\frac{\beta}{2\beta+d}}$ and which is optimal in minimax sense (See Definition 2). The latter result is proved under assumptions on $\sigma^2(\cdot)$ which are similar to the assumption imposed on the function $f(\cdot)$. In particular, β denotes the regularity of the function $\sigma^2(\cdot)$. The same result can be obtained for any noise variables ξ_i with continuously differentiable density known, possessing sufficiently many moments.

The situation changes dramatically when one considers the noise with discontinuous distribution density. Although, the transformation of the original multiplicative model to the additive one is still possible, in particular, the model (1.1) can be rewritten as

$$Y_i' = 2Y_i = f(X_i) + f(X_i)\, \eta_i, \;\; \eta_i = 2U_i - 1, \quad i = 1, \ldots, n,$$

the linear methods are not optimal anymore. As it is proved in Theorem 2.1 the optimal accuracy is given by $n^{-\frac{\beta}{\beta+d}}$. To achieve this rate the non-linear estimation procedure, based on locally bayesian approach, is proposed in Section 2.

Another interesting feature is the selection from a given family of estimators (cf. [2], [3]). Such selections are used for construction of data-driven (adaptive) procedures. In this context, several approaches to the selection from the family of linear estimators were recently proposed, see for instance [3], [4], [7] and the references therein. However, these methods are heavily based on the linearity property. As we already mentioned the locally bayesian estimators are non-linear and in Section 3 we propose the selection rule from this family. It requires, in particular, to develop new non-asymptotical exponential inequalities, which may have an independent interest.

Besides the theoretical interest, the multiplicative regression model is applied in various domains, in particular in the image processing, for example in the so-called *nonparametric frontier model* (cf. [1], [18]) can be considered as a particular case of the model (1.1). Indeed, the reconstruction of the regression function f can be viewed as the estimation of a production set $\mathcal{P}$. Indeed, $\forall i$, $Y_i \leq f(X_i)$, and therefore the estimation of f is reduced to finding the upper boundary of $\mathcal{P}$. In this context, one can also cite [10] dealing with the estimation of function's support. It is worth mentionning that although nonparametric estimation in the latter models is studied, the problem of adaptive estimation was not considered in the literature.

Minimax estimation. The first part of the paper is devoted to the minimax estimation over $\mathbb{H}_d(\beta, L, M, A)$. That means in particular that the parameters β, L, M and A are supposed to be known *a priori*. We find the *minimax rate of convergence* (1.3) on $\mathbb{H}_d(\beta, L, M, A)$ and we propose an optimal estimator in minimax sense (cf. Definition 2). Our first result (Theorem 2.1) in this direction consists in establishing a lower bound for maximal risk on $\mathbb{H}_d(\beta, L, M, A)$. We show that for any $\beta \in \mathbb{R}_+^*$, the minimax rate of convergence is bounded by the sequence

$$\varphi_n(\beta) = n^{-\frac{\beta}{\beta+d}}. \tag{1.3}$$

Next, we propose the minimax estimator, i.e. the estimator achieving the normalizing sequence (1.3). To construct the minimax estimator we use so-called *locally bayesian estimation construction* which is described in the following. Let for any $y = (y_1, \ldots, y_d) \in [0,1]^d$

$$V_h(y) = \bigotimes_{j=1}^{d} \big[y_j - h/2, y_j + h/2\big],$$

be the neighborhood around y such that $V_h(y) \subseteq [0,1]^d$, where $h \in (0,1)$ is a given scalar. Let $\mathcal{P}_b = \big\{p = (p_1, \ldots, p_d) \in \mathbb{N}^d \ : \ 0 \leq |p| \leq b\big\}$, with $|p| = p_1 + \cdots + p_d$, b is a fixed integer number and we denote D_b the cardianal of $\mathcal{P}_b$. We define the local polynomial

$$f_t(x) \quad = \quad \sum_{p \in \mathcal{P}_b} t_p \left(\frac{x-y}{h}\right)^p \mathbb{I}_{V_h(y)}(x), \quad x \in \mathbb{R}^d, \ t = (t_p \ : \ p \in \mathcal{P}_b), \tag{1.4}$$

where $z^p = z_1^{p_1} \cdots z_d^{p_d}$ for $z = (z_1, \ldots, z_d)$ and $\mathbb{I}$ denotes the indicator function. The local polynomial f_t can be viewed as an approximation of the regression function f inside the neighborhood V_h and D_b as the number of coefficients of this polynomial. One introduces the following subset of $\mathbb{R}^{D_b}$

$$\Theta\big(A,M\big) = \left\{ t \in \mathbb{R}^{D_b} : \; A \leq \sum_{p\in\mathcal{P}_b} t_p z^p \leq 3M, \; \forall z \in [-1/2,1/2]^d \right\}. \tag{1.5}$$

$\Theta(A,M)$ can be interpreted as the set of coefficients t such that $A \leq f_t(x) \leq 3M$ for all $t \in \Theta(A,M)$ and for all x in the neighbourhood $V_h(y)$. Consider the *pseudo likelihood ratio*

$$L_h\big(t, Y^{(n)}\big) = \prod_{i:\, X_i \in V_h(y)} \big[f_t(X_i)\big]^{-1} \mathbb{I}_{\left[0, f_t(X_i)\right]}\big(Y_i\big), \quad t \in \Theta\big(A,M\big).$$

Set also

$$\pi_h(t) = \int_{\Theta(A,M)} \|t-u\|_1 \, L_h\big(u, Y^{(n)}\big) du, \quad t \in \Theta\big(A,M\big), \tag{1.6}$$

where $\|\cdot\|_1$ is the ℓ_1-norm on $\mathbb{R}^{D_b}$. Let $\hat{\theta}(h)$ be the solution of the following minimization problem:

$$\hat{\theta}(h) = \arg\min_{t\in\Theta(A,M)} \pi_h(t). \tag{1.7}$$

The *locally bayesian estimator* $\bar{f}^h(y)$ of $f(y)$ is defined now as $\bar{f}^h(y) = \hat{\theta}_{0,\ldots,0}(h)$. Note that this local approach allows us to estimate successive derivatives of function f. In this paper, only the estimation of f at a given point is studied.

We note that similar locally parametric approach based on maximum likelihood estimators has been recently proposed in [8] and [17] for *regular statistical models*. But when the density of observations is discontinuous, the bayesian approach outperforms the maximum likelihood estimator. This phenomenon is well known in parametric estimation (cf. [5]). Moreover, in order to establish statistical properties of bayesian estimators, we need much weaker assumptions than whose used for analysis of maximum likelihood estimators.

As we see our construction contains an extra-parameter h to be chosen. To make this choice we use quite standard arguments. First, we note that in view of the definition of Hölder class $\mathbb{H}_d(\beta,L,M)$, we have $\forall f \in \mathbb{H}_d(\beta,L,M)$, $\exists \theta = \theta(f,y,h) \in [-M,M]^{D_b}$ such that

$$0 \leq f_\theta(x) - f(x) \leq 2Ldh^\beta, \quad \forall x \in V_h(y).$$

We will define θ in (5.5) and we will show that if $f \in \mathbb{H}_d(\beta,L,M,A)$, then $\theta \in \Theta(A,M)$ (cf. (5.6) and (5.7) below). Thus, if h is chosen sufficiently small, our original model (1.1) is well approximated inside $V_h(y)$ by the "parametric" model

$$\mathcal{Y}_i = f_\theta(X_i) \times U_i, \quad i = 1,\ldots,nh^d, \quad nh^d \in \mathbb{N}^*.$$

With this model, the *bayesian estimator* $\hat{\theta}$ is rate-optimal (See Theorem 2.2).

It is worth mentioning that the analysis of the deviation of $(X_i, \mathcal{Y}_i)_{i=1,\ldots nh^d}$ from $Y^{(nh^d)}$ is not simple. Namely here the requirements $0 < A \leq f(\cdot) \leq M$, are used. This assumption, which seems not to be necessary, allows us to make a presentation of basic ideas and to simplify routine computations (cf. also Remark 1).

Finally, $h = h_n(\beta, L) = (Ldn)^{-1/(\beta+d)}$ is chosen as the solution of the following minimization problem

$$Ldh^\beta + 1/nh^d \to \min_h . \tag{1.8}$$

Moreover we show that the corresponding estimator $\bar{f}^{h_n(\beta,L)}(y)$ is minimax for $f(y)$ on $\mathbb{H}_d(\beta, L, M, A)$ for any given value of the parameter $\beta > 0$ (cf. Theorem 2.2).

We notice that in regular statistical models where linear methods are usually optimal, the choice of the bandwidth h is due to the relation

$$Ldh^\beta + 1/\sqrt{nh^d} \to \min_h,$$

with the solution $h_{lin} = (Ldn)^{-1/(2\beta+d)}$. That explains the improvement of the rate of convergence, $n^{-\frac{\beta}{\beta+d}}$ compared to $n^{-\frac{\beta}{2\beta+d}}$, in the model with the discontinuous density.

Adaptive estimation. The second part of the paper is devoted to the adaptive minimax estimation over collection of isotropic functional classes in the model (1.1). To the best of my knowledge, the problem of adaptive estimation in the multiplicative regression with noise having discontinuous density, is not studied in the literature.

Well-known drawback of minimax approach is the dependence of the minimax estimator on the parameters describing functional class on which the maximal risk is determined. In particular, the locally bayesian estimator $\bar{f}^h(\cdot)$ depends obviously on the parameters A and M via the solution of the minimization problem (1.7). Moreover $h_n(\beta, L)$ optimally chosen in view of (1.8) depends explicitly on β and L. To overcome this drawback the minimax adaptive approach has been proposed (cf. [11], [12] and [15]). The first question arising in the adaptation can be formulated as follows.

Does there exist an estimator which would be minimax on $\mathbb{H}_d(\beta, L, M, A)$ simultaneously for all values of β, L, A and M belonging to some given subset of $\mathbb{R}_+^4$?

In section 3, we show that the answer to this question is negative, that is typical for the estimation of the function at a given point (cf. [14], [19], [20]). This answer can be reformulated in the following manner: the family of rates of convergence $\{\varphi_n(\beta),\ \beta \in \mathbb{R}_+^*\}$ is unattainable for the considered problem.

Thus, we need to find another family of normalizations for maximal risk which would be attainable and, moreover, optimal in view of some criterion of optimality. Nowadays, the most developed criterion of optimality is due to *Klutchnikoff* [9].

We show that the family of normalizations being optimal in view of this criterion, is

$$\phi_n(\beta) = \left(\frac{\rho_n(\beta)}{n}\right)^{\frac{\beta}{\beta+d}}, \qquad \rho_n(\beta) = 1 + \ln\left(\frac{\varphi_n(\beta)}{\varphi_n(b)}\right), \tag{1.9}$$

whenever $\beta \in]0, b]$ and the parameter $b > 0$ can be chosen arbitrarily. The factor ρ_n can be considered as the *price to pay for adaptation* in the context of pointwise estimation (cf. [11]).

The most important step in proving the optimality of the family (1.9) is to find an estimator, called *adaptive*, which attains the optimal family of normalizations. Obviously, we seek an estimator whose construction is *parameter-free*, i.e. independent of β, L, A and M. In order to explain our estimation procedure let us make several remarks.

First we note that the role of the constants A, M and β, L in the construction of the minimax estimator is quite different. Indeed, the constants A, M are used in order to determine the set $\Theta\big(A, M\big)$ needed for the construction of the locally bayesian estimator (see (1.6) and (1.7)). However, this set does not depend on the localization parameter $h > 0$, in other words, the quantities A and M are not involved in the selection of optimal size of the local neighborhood given by (1.8). Contrarily the constants β, L are used for the derivation of the optimal size of the local neighborhood (1.8), but they are not involved in the construction of the collection of locally bayesian estimators $\big\{\bar{f}^h, h > 0\big\}$.

We will explain how to replace the unknown quantities A and M in the definition of $\Theta\big(A, M\big)$. Indeed, a simple observation consists in the following: the estimator $\bar{f}^{h_n(\beta,L)}$ remains minimax if we replace $\Theta\big(A, M\big)$ in (1.6) and (1.7) by $\Theta\big(\tilde{A}, \tilde{M}\big)$ with any $0 < \tilde{A} \leq A$ and $M \leq \tilde{M} < \infty$. It follows from obvious inclusion $\mathbb{H}_d(\beta, L, A, M) \subseteq \mathbb{H}_d(\beta, L, \tilde{A}, \tilde{M})$. The next observation is less trivial and it follows from Proposition 1. Set $h_{\max} = n^{-\frac{1}{b+d}}$ and define for any function f

$$A(f) = \inf_{x \in V_{h_{\max}}(y)} f(x), \quad M(f) = \sum_{p \in \mathcal{P}_b} \left| \frac{\partial^m f(y)}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}} \right|. \tag{1.10}$$

The following agreement will be used in the sequel: if the function f and $m \geq 1$ are such that $\partial^m f$ does not exist we will put formally $\partial^m f = 0$ in the definition of $M(f)$.

It remains to note that contrary to the quantities A and M the functionals $A(f)$ and $M(f)$ can be consistently estimated from the observation (1.1) and one defines $\hat{A}$ and $\hat{M}$ be the corresponding estimators. Now we want to determine the collection of locally bayesian estimators $\{\hat{f}^h, h > 0\}$ by replacing $\Theta(A, M)$ in (1.6) and (1.7) by the *random* parameter set $\hat{\Theta}$ which is defined as follows.

$$\hat{\Theta} = \Theta(\hat{A}/2, 2\hat{M}) = \left\{ t \in \mathbb{R}^{D_b} : \; \hat{A}/2 \le \sum_{p \in \mathcal{P}_b} t_p z^p \le 6\hat{M}, \; \forall z \in [-1/2, 1/2]^d \right\}.$$

In this context, it is important to emphasize that the estimators $\hat{A}$ and $\hat{M}$ are built from the same observation as the one used for the construction of the family $\{\hat{f}^h, h > 0\}$.

Contrary to all saying above, the constants β and L cannot be estimated consistently. In order to select an "optimal" estimator from the family $\{\hat{f}^h, h > 0\}$ we use the general adaptation scheme due to *Lepski* [11] and [13]. To the best of our knowledge it is the first time this method is applied in the statistical model with multiplicative noise and discontinuous distribution. Moreover, except already mentioned papers [8] and [17], Lepski's procedure is typically applied to the selection from the collection of linear estimators (kernel estimators, locally polynomial estimator, etc.). In the present paper we apply this method to very complicated family of nonlinear estimators, obtained by the use of bayesian approach on the random parameter set. It requires in particular to establish the exponential inequality for the deviation of locally bayesian estimator from the parameter to be estimated (Proposition 1). It generalizes the inequality proved for the parametric model (cf. [5] Chapter 1, Section 5), this result seems to be new.

Simulations. In the present paper we adapt the local parametric approximation to a purely non parametric model. As proven, this strategy leads to the theoretically optimal statistical decisions. But the minimax as well as the minimax adaptive approach are asymptotical and it seems natural to check how proposed estimators work for reasonable sample size. In the simulation study, we test the bayesian estimator in the parametric and nonparametric cases. We show that the *adaptive* estimator approaches the *oracle* estimator. The *oracle* estimator is selected from the family $\left\{\hat{f}^h,\, h > 0\right\}$ under the hypothesis that f is known. We show that the bayesian estimator performs well starting with $n \ge 100$.

This paper is organized as follows. In Section 2 we present the results concerning minimax estimation and Section 3 is devoted to the adaptive estimation. The simulations are given in Section 4. The proofs of main results are in Section 5 (upper bounds) and section 6 (lower bounds). Auxiliary lemmas are proven in the appendix (Section 7) because they are technical results.

2. Minimax estimation on isotropic Hölder class

In this section we present several results concerning minimax estimation. First, we establish a lower bound for minimax risk defined on $\mathbb{H}_d(\beta, L, M, A)$ for any β, L, M and A.

Definition 1. *Fix $\beta > 0$, $L > 0$ and $M > 0$ and let $\lfloor \beta \rfloor$ be the largest integer strictly less than β. The* isotropic Hölder class *$\mathbb{H}_d(\beta, L, M)$ is the set of functions $f : [0,1]^d \to \mathbb{R}$ admitting on $[0,1]^d$ all partial derivatives of order $\lfloor \beta \rfloor$ and such that $\forall x, y \in [0,1]^d$*

$$\left| f(x) - \sum_{0 \le |p| \le \lfloor \beta \rfloor} \frac{\partial^{|p|} f(y)}{\partial y_1^{p_1} \cdots \partial y_d^{p_d}} \prod_{j=1}^{d} \frac{(x_j - y_j)^{p_j}}{p_j!} \right| \le L \sum_{j=1}^{d} |x_j - y_j|^\beta,$$

$$\sum_{0 \le |p| \le \lfloor \beta \rfloor} \sup_{x \in [0,1]^d} \left| \frac{\partial^{|p|} f(x)}{\partial x_1^{p_1} \cdots \partial x_d^{p_d}} \right| \le M,$$

where x_j and y_j are the j^{th} components of x and y.

This definition implies that if $f \in \mathbb{H}_d(\beta, L, M, A)$ (defined in the beginning of this paper), then $A \le A(f)$ and $M(f) \le M$, where $A(f)$ and $M(f)$ are defined in (1.10).

Maximal and minimax risk on $\mathbb{H}_d(\beta, L, M, A)$. To measure the performance of estimation procedures on $\mathbb{H}_d(\beta, L, M, A)$ we will use minimax approach.

Let $\mathbb{E}_f = \mathbb{E}_f^n$ be the mathematical expectation with respect to the probability law of the observation $Y^{(n)}$ satisfying (1.1). We define first the maximal risk on $\mathbb{H}_d(\beta, L, M, A)$ corresponding to the estimation of the function f at a given point $y \in [0,1]^d$.
Let $\tilde{f}$ be an arbitrary estimator built from the observation $Y^{(n)}$. For any $q > 0$, let

$$R_{n,q}\big[\tilde{f}, \mathbb{H}_d(\beta, L, M, A)\big] = \sup_{f \in \mathbb{H}_d(\beta, L, M, A)} \mathbb{E}_f \big|\tilde{f}(y) - f(y)\big|^q.$$

The quantity $R_{n,q}\big[\tilde{f}, \mathbb{H}_d(\beta, L, M, A)\big]$ is called *maximal risk* of the estimator $\tilde{f}$ on $\mathbb{H}_d(\beta, L, M, A)$ and the *minimax risk* on $\mathbb{H}_d(\beta, L, M, A)$ is defined as

$$R_{n,q}\big[\mathbb{H}_d(\beta, L, M, A)\big] = \inf_{\tilde{f}} R_{n,q}\big[\tilde{f}, \mathbb{H}_d(\beta, L, M, A)\big],$$

where the infimum is taken over the set of all estimators.

Definition 2. *The normalizing sequence ψ_n is called minimax rate of convergence (MRT) and the estimator $\hat{f}$ is called minimax (asymptotically minimax)* *if*

$$\liminf_{n\to\infty} \psi_n^{-q}\, R_{n,q}\big[\hat{f}, \mathbb{H}_d(\beta,L,M,A)\big] \quad > \quad 0;$$
$$\limsup_{n\to\infty} \psi_n^{-q}\, R_{n,q}\big[\hat{f}, \mathbb{H}_d(\beta,L,M,A)\big] \quad < \quad \infty.$$

Theorem 2.1. *For any $\beta > 0$, $L > 0$, $M > 0$, $A > 0$, $q \geq 1$ and $d \geq 1$*

$$\liminf_{n\to\infty} \varphi_n^{-q}(\beta) R_{n,q}\big[\mathbb{H}_d(\beta,L,M,A)\big] > 0, \qquad \varphi_n(\beta) = n^{-\frac{\beta}{\beta+d}}.$$

Remark 1. *The obtained result shows that on $\mathbb{H}_d\big(\beta,L,M,A\big)$ the minimax rate of convergence cannot be faster than $n^{-\frac{\beta}{\beta+d}}$. In view of the obvious inclusion $\mathbb{H}_d(\beta,L,M,A) \subset \mathbb{H}_d(\beta,L,M)$ the minimax rate of convergence on an isotropic Hölder class is also bounded by $n^{-\frac{\beta}{\beta+d}}$.*

Put $\bar{h} = (Ln)^{-\frac{1}{\beta+d}}$ and let $\bar{f}^{\bar{h}}(y) = \hat{\theta}_{0,\dots,0}\big(\bar{h}\big)$ is given by (1.5), (1.6) and (1.7) with $h = \bar{h}$. The next theorem shows that this estimator, based on locally bayesian approach, is minimax over Hölder classes.

Theorem 2.2. *Let $\beta > 0$, $L > 0$, $M > 0$ and $A > 0$ fixed. Then there exists a constant C^* such that for any $n \in \mathbb{N}^*$ satisfying $n\bar{h}^d \geq \big(\lfloor\beta\rfloor + 1\big)^d$*

$$\varphi_n^{-q}(\beta)\, R_{n,q}\Big[\bar{f}^{\bar{h}}, \mathbb{H}_d(\beta,L,M,A)\Big] \leq C^*, \quad \forall q \geq 1.$$

The explicit form of C^* is given in the proof.

Remark 2. *We deduce from Theorems 2.1 and 2.2 that the estimator $\bar{f}^{\bar{h}}(y)$ is minimax on $\mathbb{H}_d(\beta,L,M,A)$.*

3. Adaptive estimation on isotropic Hölder classes

This section is devoted to the adaptive estimation over the collection of the classes $\Big\{\mathbb{H}_d(\beta,L,M,A)\Big\}_{\beta,L,M,A}$. We will not impose any restriction on the possible values of L, M, A, but we will assume that $\beta \in (0,b]$, where b, as previously, is an arbitrary chosen integer.

We start by showing that there is no optimally adaptive estimator (here we follow the terminology introduced in [11], [13]). It means that there is no estimator which would be minimax simultaneously for several values of parameter β even if all other parameters L, M and A are supposed to be fixed. This result does not require any restriction on β as well.

Theorem 3.1. *For any $\mathbb{B} \subseteq \mathbb{R}^+ \setminus \{0\}$ such that card$(\mathbb{B}) \geq 2$, for any $\beta_1, \beta_2 \in \mathbb{B}$ and any $L > 0$, $M > 0$, $A > 0$*

$$\liminf_{n\to\infty} \inf_{\tilde{f}} \Big[\varphi_n^{-q}(\beta_1)\, R_{n,q}\big(\tilde{f}, \mathbb{H}_d(\beta_1,L,M,A)\big) + \varphi_n^{-q}(\beta_2)\, R_{n,q}\big(\tilde{f}, \mathbb{H}_d(\beta_2,L,M,A)\big)\Big] = +\infty,$$

where the infimum is taken over all possible estimators.

The assertion of Theorem 3.1 can be considerably specified if $\mathbb{B} = (0, b]$. To do that we need the following definition. Let $\Psi = \{\psi_n(\beta)\}_{\beta\in(0,b]}$ be a given family of normalizations.

Definition 3. *The family Ψ is called admissible if there exist an estimator $\hat{f}_n$ such that for* some *$L > 0, M > 0$ and $A > 0$*

$$\limsup_{n\to\infty} \psi_n^{-q}(\beta)\, R_{n,q}\big(\hat{f}_n, \mathbb{H}_d(\beta, L, M, A)\big) < \infty,\ \ \forall\beta \in (0, b]. \tag{3.1}$$

The estimator $\hat{f}_n$ satisfying (3.1) is called Ψ-attainable. *The estimator $\hat{f}_n$ is called* Ψ-adaptive *if (3.1) holds for* any *$L > 0, M > 0$ and $A > 0$.*

Note that Theorem 3.1 means that the family of rates of convergence $\{\varphi_n(\beta)\}_{\beta\in(0,b]}$ is not admissible.

Denote by Φ the following family of normalizations:

$$\phi_n(\beta) = \left(\frac{\rho_n(\beta)}{n}\right)^{\frac{\beta}{\beta+d}}, \quad \rho_n(\beta) = 1 + \ln\left(\frac{\varphi_n(\beta)}{\varphi_n(b)}\right),\ \ \beta \in (0, b].$$

We notice that $\phi_n(b) = \varphi_n(b)$ and $\rho_n(\beta) \sim \ln n$ for any $\beta \neq b$.

Theorem 3.2. *Let $\Psi = \{\psi_n(\beta)\}_{\beta\in(0,b]}$ be an arbitrary admissible family of normalizations.*
I. *For any $\alpha \in (0, b]$ such that $\psi_n(\alpha) \neq \varphi_n(\alpha)$, there exists an admissible family $\{\upsilon_n(\beta)\}_{\beta\in(0,b]}$ for which*

$$\lim_{n\to\infty} \upsilon_n(\alpha)\psi_n^{-1}(\alpha) = 0.$$

II. *If there exists $\gamma \in (0, b)$ such that*

$$\lim_{n\to\infty} \psi_n(\gamma)\phi_n^{-1}(\gamma) = 0, \tag{3.2}$$

then necessarily

$$(\mathbf{a}) \qquad \lim_{n\to\infty} \psi_n(\beta)\phi_n^{-1}(\beta) > 0, \quad \forall\beta \in (0, \gamma);$$

$$(\mathbf{b}) \qquad \lim_{n\to\infty} \left[\frac{\psi_n(\gamma)}{\phi_n(\gamma)}\right]\left[\frac{\psi_n(\beta)}{\phi_n(\beta)}\right] = \infty, \quad \forall\beta \in (\gamma, b].$$

We make several remarks.

We note that if the family of normalizations Φ is admissible, i.e. one can construct Φ-attainable estimator, then Φ is an *optimal* family of normalizations in view of Kluchnikoff criterion [9]. It follows from the second assertion of the theorem. We note however that a Φ-attainable estimator may depend on $L > 0, M > 0$ and $A > 0$, and therefore, this estimator have only theoretical interest. In the next section we construct a Φ-adaptive estimator which is fully parameter-free by definition. Moreover, this estimator obviously proves that Φ is admissible, and therefore optimal as it was mentioned above.

The assertions of Theorem 3.2 allows us to give rather simple interpretation of Kluchnikoff criterion. Indeed, the first assertion which is easily deduced from Theorem 3.1, shows that any admissible family of normalizations can be improved by another admissible family at any given point $\alpha \in (0, b]$ except maybe one. In particular, it concerns the family Φ if it is admissible. On the other hand, the second assertion of the theorem shows that there is no admissible family which would outperform the family Φ at two points. Moreover, in view of (**a**), Φ-adaptive (attainable) estimator, when it exists, has the same precision on $\mathbb{H}_d(\beta, L, M, A)$, $\beta < \gamma$, as any Ψ-adaptive (attainable) estimator whenever Ψ satisfies (3.2). Additionally, (**b**) implies that the gain in the precision provided by Ψ-adaptive (attainable) estimator on $\mathbb{H}_d(\gamma, L, M, A)$ leads automatically to much more losses on $\mathbb{H}_d(\beta, L, M, A)$ for any $\beta > \gamma$ compared to the precision provided by Φ-adaptive (attainable) estimator. We conclude that Φ-adaptive (attainable) estimator outperforms any Ψ-adaptive (attainable) estimator whenever Ψ satisfies (3.2). It remains to note that any admissible family not satisfying (3.2) is asymptotically equivalent to Φ.

Construction of Φ-adaptive estimator. As it was already mentioned in the introduction the construction of our estimation procedure is decomposed in several steps. First, we determine the set $\hat{\Theta}$, built from observation, which is used later in order to define the family of locally bayesian estimators. Next, based on Lepski's method (cf. [11] and [15]), we propose data-driven selection from this family.

First step: Determination of parameter set. Put $h_{\max} = n^{-\frac{1}{b+d}}$ and let $\tilde{\theta}$ be the solution of the following minimization problem.

$$\inf_{t \in \mathbb{R}^{D_b}} \sum_{i: X_i \in V_{\max}(y)}^{n} \left[2Y_i - t\, K^\top \left(\frac{X_i - y}{h_{\max}} \right) \right]^2, \quad V_{\max}(y) = V_{h_{\max}}(y),$$

where $K(z) = (z^p : p \in \mathcal{P}_b)$ is the D_b-dimensional vector of monomials and the sign $\top$ means the transposition. Thus, $\tilde{\theta}$ is the local least squared estimator and its explicit expression is given by

$$\tilde{\theta} = 2 \left[\sum_{i: X_i \in V_{\max}(y)}^{n} K^\top \left(\frac{X_i - y}{h_{\max}} \right) K \left(\frac{X_i - y}{h_{\max}} \right) \right]^{-1} \left[\mathcal{K}_n(y) \right]^\top Y,$$

where $Y = (Y_1, \ldots, Y_n)$ and $\mathcal{K}_n(y) = \left[K^\top \left(\frac{X_i - y}{h_{\max}} \right) \mathbb{I}_{V_{\max}(y)}(X_i) \right]_{i=1,\ldots n}$ is the design matrix. Put

$$\tilde{\delta}_p = p_1!...p_d!\, h_{\max}^{-|p|}\, \tilde{\theta}_p, \quad |p| \le b.$$

We introduce the following quantities

$$\hat{A} = \tilde{\delta}_{0\ldots,0}, \qquad \hat{M} = \left\| \tilde{\delta} \right\|_1, \tag{3.3}$$

and we recall the definition of random parameter set

$$\hat{\Theta} = \left\{ t \in \mathbb{R}^{D_b} : \ \hat{A}/2 \leq \sum_{p \in \mathcal{P}_b} t_p z^p \leq 6\hat{M}, \ \forall z \in [-1/2, 1/2]^d \right\}. \tag{3.4}$$

<u>*Second step: Collection of locally bayesian estimators.*</u> Put

$$\hat{\pi}_h(t) = \int_{\hat{\Theta}} \|t-u\|_1 \, L_h\big(u, Y^{(n)}\big) du; \tag{3.5}$$

$$\hat{\theta}^*(h) = \arg\min_{t \in \hat{\Theta}} \hat{\pi}_h(t). \tag{3.6}$$

The family of locally bayesian estimator $\hat{\mathcal{F}}$ is defined now as follows.

$$\hat{\mathcal{F}} = \left\{ \hat{f}^h(y) = \hat{\theta}^*_{0,\dots,0}(h), \ h \in \big(0, h_{\max}\big] \right\}. \tag{3.7}$$

<u>*Third step: Data-driven selection from the collection $\hat{\mathcal{F}}$.*</u> Put

$$h_k = 2^{-k} h_{\max}, \ \ k = 0, \dots, \mathrm{k}_n,$$

where k_n is the largest integer such that $h_{\mathrm{k}_n} \geq h_{\min} = \ln^{\frac{b}{d(b+d)}} n^{-1/d}$. Set

$$\hat{\mathcal{F}}^* = \left\{ \hat{f}^{(k)}(y) = \hat{\theta}^*_{0,\dots,0}(h_k), \ \ k = 0, \dots, \mathrm{k}_n \right\}.$$

We put $\hat{f}^*(y) = \hat{f}^{(\hat{k})}(y)$, where $\hat{f}^{(\hat{k})}(y)$ is selected from $\hat{\mathcal{F}}^*$ in accordance with the rule:

$$\hat{k} = \inf \left\{ k = \overline{0, \mathrm{k}_n} : \ \ \big|\hat{f}^{(k)}(y) - \hat{f}^{(l)}(y)\big| \leq \hat{M} S_n\big(l\big), \ \ l = \overline{k+1, \mathrm{k}_n} \right\}. \tag{3.8}$$

Here we have used the following notations.

$$S_n(l) = 432 D_b^3 (32qd + 16) \, \lambda_n^{-1}\big(h_l\big) \, \left[\frac{1 + l \ln 2}{n\big(h_l\big)^d} \right], \quad l = 0, 1, \dots, \mathrm{k}_n,$$

and $\lambda_n(h)$ is the smallest eigenvalue of the matrix

$$\mathcal{M}_{nh}(y) = \frac{1}{nh^d} \sum_{i=1}^n K^\top \left(\frac{X_i - y}{h} \right) K \left(\frac{X_i - y}{h} \right) \mathbb{I}_{V_h(y)}(X_i), \tag{3.9}$$

which is completely determined by the design points and by the number of observations. We will prove that there exists a nonnegative real λ, such that $\lambda_n(h) \geq \lambda$ for any $n \geq 1$ and any $h \in \big[h_{\min}, h_{\max}\big]$ (cf. Lemma 2).

Theorem 3.3. *Let an integer number $b > 0$ fixed. Then for any $\beta \in (0, b]$, $L > 0, \ M > 0, A > 0$ and $q \geq 1$*

$$\limsup_{n \to \infty} \phi_n^{-q}(\beta) \, R_{n,q}\Big[\hat{f}^*, \mathbb{H}_d(\beta, L, M, A)\Big] < \infty.$$

Remark 3. *The assertion of the theorem means that the proposed estimator $\hat{f}^*(y)$ is Φ-adaptive. It implies in particular that the family of normalizations Φ is admissible. This, together with Theorem 3.2 allows us to state the optimality of Φ in view of Kluchnikoff criterion (cf. [9]).*

4. Simulation study

We will consider the case $d = 1$. The data are simulated accordingly to the model (1.1), where we use the following functions (Figure 1).

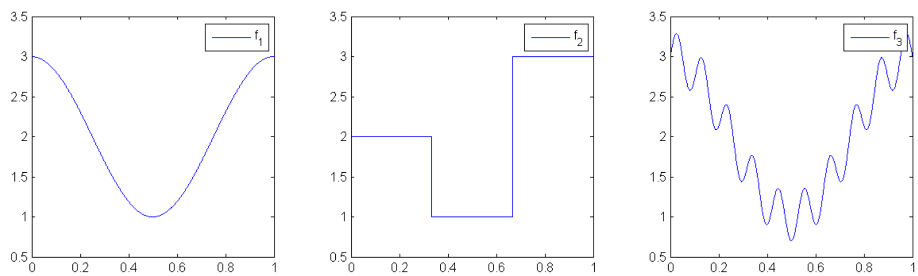

FIGURE 1. *Test functions.*

Here $f_1(x) = \cos(2\pi x) + 2$, $f_2(x) = 2.\mathbb{I}_{[x\leq 1/3]} + 1.\mathbb{I}_{[1/3<x\leq 2/3]} + 3.\mathbb{I}_{[2/3<x]}$ and $f_3(x) = \cos(2\pi x) + 2 + 0.3\sin(19\pi x)$

To construct the family of estimators we use the linear approximation ($b = 2$), i.e. within the neighbourhoods of the given size h, the locally bayesian estimator has the form

$$\hat{f}^h(x) = \hat{\theta}_0 + \hat{\theta}_1 x, \quad x \in [0, 1].$$

We define the ideal (oracle) value of the parameter $\tilde{h} = \tilde{h}(f)$ as the minimizer of the risk:

$$\tilde{h} = \arg \inf_{h\in[1/n,1]} \mathbb{E}_f \left|\hat{f}^h(y) - f(y)\right|.$$

To compute it we apply Monte-Carlo simulations (10000 repetitions). Our first objective is to compare the risk provided by the "oracle" estimator $\hat{f}^{\tilde{h}}(\cdot)$ and the one provided by the adaptive estimator from Section 3. Figure 2 shows the deviation of the adaptive estimator from the function to be estimated. In several points, for example in $y = 1/2$, we notice the so-called over-smoothing phenomenon, inherent to any adaptive estimator.

Oracle-adaptive ratio. We compute the risks of the oracle and the adaptive estimator in 100 points of the interval $(0, 1)$. The next tabular presents the mean value of the ratio oracle risk/adaptive risk calculated for the functions f_1, f_2, f_3 and $n = 100, 1000$.

Figure 4 presents the "oracle risk/adaptive risk" ratio as the function of the number of observations n.

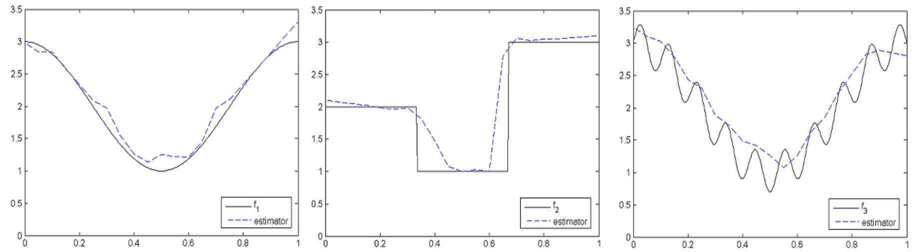

FIGURE 2. *Examples of estimation with $n = 100$.*

	n = 100		n = 1000	
function	adaptive risk	oracle-adaptive ratio	adaptive risk	oracle-adaptive ratio
f_1	0.13	0.84	0.03	0.85
f_2	0.3	0.71	0.1	0.75
f_3	0.28	0.65	0.2	0.68

FIGURE 3. *Numeric values of risk.*

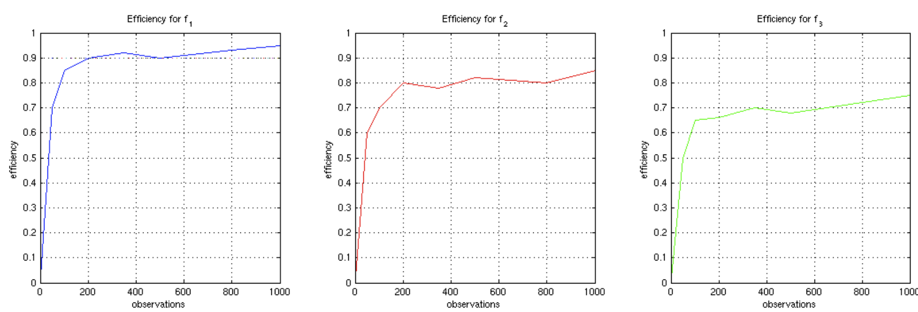

FIGURE 4. *Efficiency of bayesian estimator for three test functions.*

Adaptation versus parametric estimation. We consider the function f_4 (figure 5), which is linear inside the neighborhood of size $h_* = 1/8$ around point $1/2$ and simulate $n = 1000$ observations in accordance with the model (1.1). Using only the observations corresponding to the interval $[3/8, 5/8]$ we construct the bayesian estimator $\hat{f}^{1/8}(1/2)$.

It is important to emphasize that this estimator is efficient [5] since the model is parametric. Now our objective is to compare the risk of our adaptive estimator with the risk provided by the estimator $\hat{f}^{1/8}(1/2)$. We also try to understand how far is the localization parameter $h_{\hat{k}}$ from the true value $1/8$, inherent to the construction of our adaptive estimator. We compute the risk of each estimator via Monte-Carlo method with 10000 repetitions. For each repetition the procedure select the adaptive bandwidth $h_{\hat{k}}^{(j)}$, $j = 1, ..., 10000$.

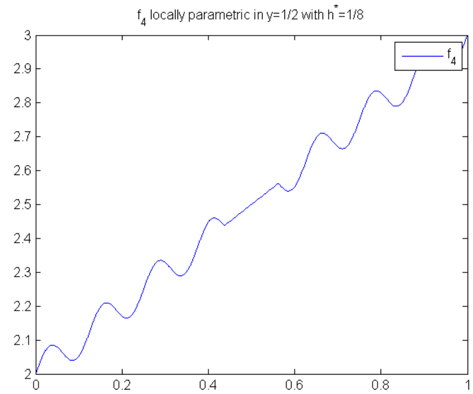

FIGURE 5. *local parametric test function.*

We confirm once again the over-smoothing phenomenon since

$$h^{(j)}_{\hat{k}} \sim 0.1405 > h_* = 0.1250, \quad j = 1, ..., 10000.$$

Note however that the adaptive procedure selects the size of the neighborhood which is quite close to the true one. We also compute the risks of both estimators: "bayesian risk"=**0.0206** and "adaptive risk"=**0.0308**. We conclude that the estimation accuracy provided by our adaptive procedure is quite satisfactory.

5. Proofs of main results: upper bounds

Let $\mathcal{H}_n, n > 1$ be the following subinterval of $(0, 1)$.

$$\mathcal{H}_n = \left[\frac{\big(b+1\big) \vee \big(\ln n\big)^{\frac{1}{(d+d^2)}}}{n^{1/d}}, \left(\frac{1}{\ln n}\right)^{\frac{1}{b+d}} \right]. \tag{5.1}$$

Later on we will consider only the values of h belonging to $\mathcal{H}_n$. We start with establishing the exponential inequality for the deviation of locally bayesian estimator $\hat{f}^h(y)$ from $f(y)$. The corresponding inequality is the basic technic to prove minimax and minimax adaptive results.

5.1. Exponential Inequality

Let us introduce the following notations. For any $h \in \mathcal{H}_n$, put $\omega = \omega(f, y, h) = \big\{\omega_p : p \in \mathcal{P}_b\big\}$, where $\omega_0 = \omega_{0,...,0} = f(y)$ and

$$\omega_p = \frac{\partial^{|p|} f(y)}{\partial y_1^{p_1} \cdots \partial y_d^{p_d}} \frac{h^{|p|}}{p_1!...p_d!}, \quad p \neq 0. \tag{5.2}$$

Remind the agreement we follow in the present paper: if the function f and vector p are such that $\partial^{|p|}f$ does not exist we put $\omega_p = 0$.

Let f_ω given by (1.4), be the local polynomial approximation of f inside $V_h(y)$ and let b_h be the corresponding approximation error, i.e.

$$b_h = \sup_{x \in V_h(y)} \big|f_\omega(x) - f(x)\big|. \tag{5.3}$$

If $f \in \mathbb{H}_d(\beta, L, M)$, $\beta > 0$, one could remark that $b_h \le Ldh^\beta$ by the definition of ω in (5.2) and Definition 2. Put also

$$\mathcal{N}_h = b_h \times nh^d, \quad \mathcal{E}(h) = \exp\left\{\frac{(1+6D_b^2)\mathcal{N}_h}{6A(f)D_b^2}\right\}. \tag{5.4}$$

Introduce the random events $G_{\hat{M}} = \big\{\big|\hat{M} - M(f)\big| \le M(f)/2\big\}$ and $G_{\hat{A}} = \big\{\big|\hat{A} - A(f)\big| \le A(f)/2\big\}$ and set $G = G_{\hat{M}} \cap G_{\hat{A}}$ where $\hat{A}$ and $\hat{M}$ are defined in (3.3).

Recall that $\lambda_n(h)$ (cf. Section 3) is the smallest eigenvalue of the matrix

$$\mathcal{M}_{nh}(y) = \frac{1}{nh^d}\sum_{i=1}^n K^\top\left(\frac{X_i - y}{h}\right) K\left(\frac{X_i - y}{h}\right) \mathbb{I}_{V_h(y)}(X_i),$$

and $K(z)$ is the D_b-dimensional vector of the monomials z^p, $p \in \mathcal{P}_b$. In the sequel, we denote $\mathbb{P}_f$ the probability measure of the observation $Y^{(n)}$.

Proposition 1. *For any $h \in \mathcal{H}_n$ and any f such that $A(f) > A$ and $M(f) < M$, then $\forall \varepsilon > 144MD_b(1 \vee \mathcal{N}_h)/A\lambda_n(h)$*

$$\mathbb{P}_f\left(nh^d\big|\hat{f}^h(y) - f(y)\big| \ge \varepsilon,\, G\right) \le \mathfrak{B}\big(A(f), M(f)\big)\mathcal{E}(h)\exp\left\{-\frac{\lambda_n(h)\,\varepsilon}{432M(f)\,D_b^3}\right\},$$

where $\hat{f}^h(y) \in \hat{\mathcal{F}}$ as defined in (3.7). The explicit expression of the function $\mathfrak{B}(\cdot,\cdot)$ is given in the beginning of the proof of the proposition.

The next proposition gives us an upper bound for the risk of a locally bayesian estimator.

Proposition 2. *For any $n \in \mathbb{N}^*$, $h \in \mathcal{H}_n$ and any $f \in \mathbb{H}_d(\beta, L, M, A)$, then $\exists \lambda > 0$ such that $\lambda_n(h) \ge \lambda$ and*

$$\mathbb{E}_f\big|\hat{f}^h(y) - f(y)\big|^q \mathbb{I}_G \le C_q^*\big(A(f), M(f)\big)\left[\frac{1 \vee Ld\,nh^{\beta+d}}{nh^d}\right]^q, \quad q \ge 1,$$

where

$$C_q^*(a,m) = \frac{1}{q}\left[\frac{432mD_b^3(1+6D_b^2)}{3\lambda a D_b^2}\right]^q + \big[864m\lambda^{-1}D_b^3\big]^q\,\mathfrak{B}(a,m)\Gamma(q), \quad a, m > 0,$$

$\Gamma(\cdot)$ is the well-known Gamma function.

Remark 4. *The analysis of the proof of Proposition 1 allows to assert the following inequality*

$$\mathbb{P}_f\left(nh^d\big|\bar{f}^h(y)-f(y)\big|\geq\varepsilon\right)\leq\mathfrak{B}\big(A,M\big)\mathcal{E}(h)\exp\left\{-\frac{\lambda_n(h)\,\varepsilon}{432M\,D_b^3}\right\},$$

where $\bar{f}^h(y)$ is the locally bayesian estimator which is the minimizer of (1.6).

Thus, the latter inequality can be viewed as an analogue of the result of Proposition 1 when A and M are known. By the same reasons, we have

$$\mathbb{E}_f\big|\bar{f}^h(y)-f(y)\big|^q\leq C_q^*(A,M)\left[\frac{1\vee Ld\,nh^{\beta+d}}{nh^d}\right]^q,\quad q\geq 1.$$

5.2. Proof of Proposition 1

Before starting the proof, let us breafly discuss its ingredients.

Discussion.

I. Remind that $\hat{\theta}^*(h)$ minimizes $\hat{\pi}_h$ as defined in (3.5). Hence, the obvious following inclusion

$$\left\{nh^d\big\|\hat{\theta}^*(h)-\theta\big\|_1\geq\varepsilon\right\}\subseteq\left\{\inf_{nh^d\|t-\theta\|_1\geq\varepsilon}\hat{\pi}_h(t)\leq\hat{\pi}_h(\theta)\right\}.$$

allows us to reduce the study of the deviation of $\hat{\theta}^*(h)$ from θ to the study of the behaviour of $\hat{\pi}_h$. Here, the vector $\theta=\theta(f,y,h)=\big\{\theta_p:\ p\in\mathcal{P}_b\big\}$ is defined as follows.

$$\theta_0=\theta_{0,\dots,0}=\omega_0+b_h,\quad \theta_p=\omega_p,\ \ |p|\neq 0, \tag{5.5}$$

where the vector ω is the coefficients of Taylor polynomial defined in (5.2). The definition of b_h in (5.3) implies trivially

$$f_\theta(x)\geq f(x)\geq A(f)\geq A,\quad \forall x\in V_h(y). \tag{5.6}$$

By definition of $\mathbb{H}_d(\beta,L,M,A)$, we can see that $\forall x\in V_h(y)$

$$f_\theta(x)=f_\omega(x)+b_h\leq 2\|\omega\|_1+\|f\|_\infty\leq 3M(f)\leq 3M, \tag{5.7}$$

where $\|\cdot\|_\infty$ is the sup-norm.

Thus we have $\theta\in\Theta\big(A(f),M(f)\big)\subseteq\Theta(A,M)$. Under the event G, we notice that $\Theta\big(A(f),M(f)\big)\subseteq\hat{\Theta}=\Theta\big(\hat{A}/2,2\hat{M}\big)\subseteq\Theta\big(A(f)/4,3M(f)\big)$ where Θ and $\hat{\Theta}$ are respectively defined in (1.5) and (3.4).

II. We note that $\hat{\pi}_h$ is the integral functional of the pseudo-likelihood L_h. As the consequence, the behaviour of $\hat{\pi}_h$ is completely determined by this process. Following [5] (Chapter 1, Section 5, Theorem 5.2), where similar problems were studied under parametric model assumption, we introduce the stochastic process

$$Z_{h,\theta}(u)=\frac{L_h\big(\theta+(nh^d)^{-1}u,Y^{(n)}\big)}{L_h\big(\theta,Y^{(n)}\big)}.$$

defined on $\Upsilon_n = \left\{u \in \mathbb{R}^{D_b} \ : \ u = nh^d(t-\theta),\ t \in \Theta\big(A(f)/4, 3M(f)\big)\right\}$.

As it was noted in [5] (Chapter 1, Section 5, Theorem 5.2) the following properties of the process $Z_{h,\theta}$ are essential for the study of $\hat{\pi}_h$:

- Hölder continuity of its trajectories;
- the rate of its decay at infinity.

The exact statements are formulated in Lemma 1 below.

III. As it was shown in [5] (Chapter 1, Section 5, Theorem 5.2) in parametric situation the properties of $Z_{h,\theta}$ mentioned above, gives desirable behavior of the following process

$$z_h(u) = \frac{Z_{h,\theta}(u)}{\int_{\hat{\Upsilon}_n} Z_{h,\theta}(v)dv}, \quad u \in \hat{\Upsilon}_n := nh^d\big(\hat{\Theta} - \theta\big),$$

where the set $\hat{\Theta}$ is defined in (3.4). The exact statements are given in Assertions 1 and 2. The latter process is important in view of the following inclusion

$$\left\{nh^d\big|\hat{f}^h(y) - f(y)\big| \geq \varepsilon\right\} \subseteq \left\{\int_{\hat{\Upsilon}_n(r)} \|u\|_1 z_h(u)du > \frac{r}{2}\right\}.$$

Auxiliary Lemma. First, we note that in view of (5.6), for any $X_i \in V_h(y)$ the event $Y_i \leq f_\theta(X_i)$ is always realized, because $Y_i \leq f(X_i) \leq f_\theta(X_i)$. Hence, $Z_{h,\theta}$ can be rewritten

$$Z_{h,\theta}(u) = \prod_{i:\, X_i \in V_h(y)} \frac{f_\theta(X_i)}{f_{\theta+u(nh^d)^{-1}}(X_i)} \mathbb{I}_{\left[Y_i \leq f_{\theta+u(nh^d)^{-1}}(X_i)\right]}, \quad u \in \Upsilon_n. \tag{5.8}$$

Lemma 1. *For any $f \in \mathbb{H}_d(\beta, L, M, A)$ and $h \in \mathcal{H}_n$*

1. $\sup_{u_1,u_2 \in \Upsilon_n} \|u_1 - u_2\|_1^{-1} \mathbb{E}_f \big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big| \leq \mathcal{C}_h$,
2. $\mathbb{E}_f Z_{h,\theta}^{1/2}(u) \leq e^{-g_h\left(\|u\|_1\right)}, \ \forall u \in \Upsilon_n$,
3. $\mathbb{P}_f\left\{\int_{[0,\delta]^{D_b}} Z_{h,\theta}(u)du < \frac{\delta^{D_b}}{2}\right\} < 2\mathcal{C}_h\delta, \ \forall \delta > 0.$

where

$$\mathcal{C}_h = 8\big(1 \vee 4D_b A^{-1}(f)\big) \exp\left\{1 + \mathcal{N}_h / A(f)\right\}, \quad g_h(a) = \frac{\lambda_n(h)a}{18M(f)D_b} - \frac{\mathcal{N}_h}{A(f)},$$

with $a > 0$ and $\lambda_n(h)$ is the smallest eigenvalue of the matrix $\mathcal{M}_{nh}(y)$ defined in (3.9).

Proof of Proposition 1. Define for any $u > 0$ and $v > 0$

$$\mathfrak{B}(a, m) = \sup_{z \geq 0} 16e\big(1 \vee D_b a^{-1}\big)\Sigma(m)\left[\mathcal{B}_z + 6\right] \exp\left\{-\frac{\lambda z}{432 v D_b^3}\right\}, \tag{5.9}$$

where $\mathcal{B}_z = z^{D_b+1} + 2(2z+2)^{2D_b} + 5 + D_b\left(z^{D_b} + (2z+2)^{\frac{D_b}{2}-1}\right)$, $\lambda > 0$ is defined such that: $\lambda_n(h) \geq \lambda$ for any $n \in \mathbb{N}^*$, $h \in \mathcal{H}_n$ (for more details, see Lemma 2) and

$$\Sigma(v) = \frac{c^2(v)\big(3 - c(v)\big)}{\big(1 - c(v)\big)^3}, \quad c(v) = \exp\big\{ -(54vD_b^2)^{-1}\big\}, \quad v > 0.$$

Assertion 1. *For any $\varepsilon > 0$, and for all r such that $0 < r < \varepsilon/3$, we assume*

$$\mathbb{P}_f\left(nh^d\big|\hat{f}^h(y) - f(y)\big| \geq \varepsilon, G\right) \leq 2\mathbb{P}_f\left(\int_{\hat{\Upsilon}_n \cap (\|u\|_1 > \varepsilon/4)} \|u\|_1 z_h(u) du > \frac{r}{2}, G\right).$$

Assertion 2. *For all $h \in \mathcal{H}_n$ and any f such that $A(f) > A$ and $M(f) < M$, then for any $a > 36M(f)D_b(1 \vee \mathcal{N}_h)/(\lambda A(f))$*

$$\mathbb{E}_f\left[\int_{\hat{\Upsilon}_n \cap \left\{\|u\|_1 > a\right\}} \|u\|_1 z_h(u)\, du\, \mathbb{I}_G\right] \leq a\Sigma\big(M(f)\big)\, \mathcal{B}_a\, \mathcal{C}_h \exp\left\{-\frac{1}{6D_b} g_h(a)\right\},$$

where $g_h(\cdot)$ is defined in Lemma 1.

1^0. Suppose that Assertions 1 and 2 are proved. Then, in view of Assertion 2, choosing $a = \varepsilon/4$, we get

$$\mathbb{E}_f \int_{\hat{\Upsilon}_n \cap (\|u\|_1 > \varepsilon/4)} \|u\|_1 z_h(u)\, \mathbb{I}_G\, du \leq \frac{\varepsilon}{4}\, \Sigma\big(M(f)\big)\, \mathcal{B}_{\varepsilon/4}\, \mathcal{C}_h\, e^{-\frac{1}{6D_b^2} g_h(\varepsilon/4)}.$$

Using the Tchebychev inequality, we have in view of the last inequality

$$\mathbb{P}_f\left(\int_{\hat{\Upsilon}_n \cap (\|u\|_1 > \varepsilon/4)} \|u\|_1 z_h(u) du > \frac{\varepsilon}{8},\, G\right) \leq 2\Sigma\big(M(f)\big)\, \mathcal{B}_{\varepsilon/4}\, \mathcal{C}_h\, e^{-\frac{1}{6D_b^2} g_h(\varepsilon/4)}.$$

The assertion of Proposition 1 follows now from the last inequality, Assertion 1 and the definitions of $\mathcal{C}_h, g_h(\cdot)$ and the function $\mathfrak{B}(\cdot,\cdot)$.

2^0. Now, we will prove Assertion 1. The definitions of $\hat{\theta}^*(h)$ and $\theta = \theta(f, y, h)$ imply $\forall \varepsilon > 0$

$$\begin{aligned} \mathbb{P}_f\left(nh^d\big|\hat{f}^h(y) - f(y)\big| \geq \varepsilon, G\right) &\leq \mathbb{P}_f\left(nh^d\big|\hat{\theta}^*_{0,\dots,0}(h) - \theta_{0,\dots,0}\big| \geq \varepsilon, G\right) \\ &\leq \mathbb{P}_f\left(nh^d\big\|\hat{\theta}^*(h) - \theta\big\|_1 \geq \varepsilon, G\right). \end{aligned} \tag{5.10}$$

We can make some remarks. First, it is easily seen that $\theta \in \Theta\big(A(f), M(f)\big)$ (cf. (5.6) and (5.7)). Therefore, if the event G holds then $\theta \in \hat{\Theta}$. Remind also that $\hat{\theta}^*(h)$ minimizes $\hat{\pi}_h$ defined in (3.5) and, therefore, the following inclusion holds since $\hat{\theta}^*(h) \in \hat{\Theta}$.

$$\left\{\left(nh^d\big\|\hat{\theta}^*(h) - \theta\big\|_1 \geq \varepsilon\right) \cap G\right\} \subseteq \left\{\left(\inf_{nh^d\|t-\theta\|_1 \geq \varepsilon} \hat{\pi}_h(t) \leq \hat{\pi}_h(\theta)\right) \cap G\right\}. \tag{5.11}$$

Moreover,

$$\begin{aligned}\hat{\pi}_h(t) &= (nh^d)^{-1}\int_{\hat{\Theta}}\left\|nh^d(t-u)\right\|_1 L_h\big(u,Y^{(n)}\big)du\\ &= (nh^d)^{-D_b-1}\int_{\hat{\Upsilon}_n}\left\|nh^d(t-\theta)-u\right\|_1 L_h\big(\theta+u(nh^d)^{-1},Y^{(n)}\big)du\\ &= (nh^d)^{-D_b-1}L_h\big(\theta,Y^{(n)}\big)\int_{\hat{\Upsilon}_n}\left\|nh^d(t-\theta)-u\right\|_1 Z_{h,\theta}(u)du.\end{aligned}$$

Hence, $\tau_n = nh^d\big(\hat{\theta}^*(h)-\theta\big)$ is the minimizer of

$$\chi_n(s)=\int_{\hat{\Upsilon}_n}\left\|s-u\right\|_1\frac{Z_{h,\theta}(u)}{\int_{\hat{\Upsilon}_n}Z_{h,\theta}(v)dv}du$$

and we obtain from (5.10) and (5.11) that for any $\varepsilon>0$

$$\mathbb{P}_f\left(\left\|nh^d(\hat{\theta}^*(h)-\theta)\right\|_1>\varepsilon,G\right)\le\mathbb{P}_f\left(\inf_{\|s\|_1>\varepsilon}\chi_n(s)\le\chi_n(0),G\right). \tag{5.12}$$

Let $0<r<\varepsilon/3$, be a number whose choice will be done later. We have

$$\chi_n(0)\le r\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}z_h(u)du+\int_{\hat{\Upsilon}_n\cap(\|u\|_1>r)}\|u\|_1 z_h(u)du.$$

Note also that

$$\begin{aligned}\inf_{\|s\|_1>\varepsilon}\chi_n(s) &\ge \inf_{\|s\|_1>\varepsilon}\left[\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}\big(\|s\|_1-\|u\|_1\big)z_h(u)du\right]\\ &\ge (\varepsilon-r)\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}z_h(u)du.\end{aligned}$$

It yields in particular that

$$\begin{aligned}&\chi_n(0)-\inf_{\|s\|_1>\varepsilon}\chi_n(s)\\ &\quad\le-(\varepsilon-2r)\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}z_h(u)du+\int_{\hat{\Upsilon}_n\cap(\|u\|_1>r)}\|u\|_1 z_h(u)du.\end{aligned}$$

Thus, $\forall r\in(0,\varepsilon/3)$

$$\begin{aligned}&\mathbb{P}_f\left(\chi_n(0)-\inf_{\|s\|_1>\varepsilon}\chi_n(s)>0,G\right)\\ &\le\mathbb{P}_f\left(\int_{\hat{\Upsilon}_n\cap(\|u\|_1>r)}\|u\|_1 z_h(u)du>(\varepsilon-2r)\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}z_h(u)du,G\right)\\ &\le\mathbb{P}_f\left(\int_{\hat{\Upsilon}_n\cap(\|u\|_1>r)}\|u\|_1 z_h(u)du>r/2,G\right)\\ &\quad+\mathbb{P}_f\left((\varepsilon-2r)\int_{\hat{\Upsilon}_n\cap(\|u\|_1\le r)}z_h(u)du<r/2,G\right).\end{aligned} \tag{5.13}$$

We note that the second term in (5.13) can be control by the first one whenever $0 < r < \varepsilon/3$. Indeed, setting $\hat{\Upsilon}_n(r) = \hat{\Upsilon}_n \cap (u \in \mathbb{R}^{D_b} \;:\; \|u\|_1 > r)$ we get

$$\begin{aligned}
&\mathbb{P}_f\left((\varepsilon - 2r)\int_{\hat{\Upsilon}_n \cap (\|u\|_1 \le r)} z_h(u)du < r/2, G\right) \\
&\le \mathbb{P}_f\left(r\int_{\hat{\Upsilon}_n} Z_{h,\theta}(v)dv - r\int_{\hat{\Upsilon}_n(r)} Z_{h,\theta}(u)du < \frac{r}{2}\int_{\hat{\Upsilon}_n} Z_{h,\theta}(v)dv, G\right) \\
&\le \mathbb{P}_f\left(r\int_{\hat{\Upsilon}_n(r)} Z_{h,\theta}(u)du > \frac{r}{2}\int_{\hat{\Upsilon}_n} Z_{h,\theta}(v)dv, G\right) \\
&\le \mathbb{P}_f\left(\int_{\hat{\Upsilon}_n(r)} \|u\|_1 z_h(u)du > r/2, G\right).
\end{aligned}$$

The last inequality together with (5.10), (5.12) and (5.13) yields

$$\mathbb{P}_f\left(nh^d\big|\hat{f}^h(y) - f(y)\big| \ge \varepsilon, G\right) \le 2\mathbb{P}_f\left(\int_{\hat{\Upsilon}_n(r)} \|u\|_1 z_h(u)du > \frac{r}{2}, G\right).$$

3^0. Now, let us prove Assertion 2. Put $\Upsilon_n(a) = \Upsilon_n \cap (u \in \mathbb{R}^{D_b} \;:\; \|u\|_1 > a)$ for all $a > 0$ and $\Omega_\upsilon = \Upsilon_n(\upsilon) \setminus \Upsilon_n(\upsilon+1)$ for any $\upsilon \ge a$. Introduce the following notations.

$$\mathcal{I}_\upsilon = \int_{\Omega_\upsilon} Z_{h,\theta}(u)du, \qquad \mathcal{Q}_\upsilon = \frac{\int_{\hat{\Upsilon}_n \cap \Omega_\upsilon} Z_{h,\theta}(u)du}{\int_{\hat{\Upsilon}_n} Z_{h,\theta}(u)du}.$$

Fix $T > 0$ whose choice will be done later. Consider the minimal number $N(\Omega_\upsilon, 1/T)$ of balls of radius $1/T$ that are needed to cover the set Ω_υ. Denote by u^j the center of each ball. Since Ω_υ is a compact of $\mathbb{R}^{D_b}$, it implies $N(\Omega_\upsilon, 1/T) \le (\upsilon+1)^{D_b} T^{D_b}$. Introduce the non-intersecting parts $\Delta_1, \Delta_2, \Delta_3, \dots$ as follows: $\Delta_1 = \left\{u \in \Omega_\upsilon \;:\; \|u - u^1\|_1 \le 1/T\right\}$ and

$$\Delta_j = \left\{u \in \Omega_\upsilon \;:\; \|u - u^j\|_1 \le 1/T\right\} \setminus \bigcup_{l=1}^{j-1} \Delta_l, \quad j = 2, \dots, N(\Omega_\upsilon, 1/T).$$

Put $S_\upsilon = \sum_j \int_{\Delta_j} Z_{h,\theta}(u^j)du$ and notice that S_υ is stepwise approximation of $\mathcal{I}_\upsilon$.

Control of $\mathcal{I}_\upsilon$. Remind that $\Omega_\upsilon = \bigcup_{j=1}^{N(\Omega_\upsilon,1/T)} \Delta_j$ and denote by $|\Omega_\upsilon|$ the volume of Ω_υ. We get for any $\sigma > 0$

$$\begin{aligned}
\mathbb{P}_f\big(S_\upsilon > \sigma\big) &\le \mathbb{P}_f\left(\max_j Z_{h,\theta}^{1/2}(u^j)\sqrt{|\Omega_\upsilon|} > \sqrt{\sigma}\right) \\
&\le \sum_j \mathbb{P}_f\left(Z_{h,\theta}^{1/2}(u^j) > \sqrt{|\Omega_\upsilon|}\sqrt{\sigma}\right).
\end{aligned}$$

Note that the number of summands on the right-hand side of the last inequality does not exceed $(\upsilon+1)^{D_b}T^{D_b}$. Applying Tchebychev inequality and Lemma 1 (*2*), we obtain

$$\mathbb{P}_f\big(S_\upsilon > \sigma\big) \le (\upsilon+1)^{D_b}T^{D_b}\sqrt{|\Omega_\upsilon|}\sigma^{-1/2}e^{-g_h(\upsilon)}. \tag{5.14}$$

In view of to Lemma 1 (*1*),

$$\mathbb{E}_f\big|S_\upsilon - \mathcal{I}_\upsilon\big| \le \sum_j \int_{\Delta_j} \mathbb{E}_f\big|Z_{h,\theta}(u) - Z_{h,\theta}(u^j)\big|du \le \mathcal{C}_h \sum_j \int_{\Delta_j} \|u-u^j\|_1 du.$$

By definition of Δ_j, each summand does not exceed $\int_{\Delta_j} T^{-1}du$, therefore,

$$\mathbb{E}_f\big|S_\upsilon - \mathcal{I}_\upsilon\big| \le \mathcal{C}_h|\Omega_\upsilon|T^{-1}. \tag{5.15}$$

One has

$$\mathbb{P}_f\big(\mathcal{I}_\upsilon > 2\sigma\big) \le \mathbb{P}_f\big(S_\upsilon > \sigma\big) + \mathbb{P}_f\big(\big|S_\upsilon - \mathcal{I}_\upsilon\big| > \sigma\big).$$

Using (5.14), (5.15) and applying Tchebychev inequality, we get

$$\mathbb{P}_f\big(\mathcal{I}_\upsilon > 2\sigma\big) \le (\upsilon+1)^{D_b}T^{D_b}\sqrt{|\Omega_\upsilon|}\sigma^{-1/2}e^{-g_h(\upsilon)} + \mathcal{C}_h|\Omega_\upsilon|T^{-1}\sigma^{-1}. \tag{5.16}$$

Control of $\mathcal{Q}_\upsilon$. Set $\mathbb{A} = \Big\{\int_{\hat{\Upsilon}_n} Z_{h,\theta}(u)du < \delta^{D_b}/2\Big\}$. Since $\mathcal{Q}_\upsilon \le 1$ we obtain for any $\delta > 0$, $\sigma > 0$

$$\begin{aligned}\mathbb{E}_f\mathcal{Q}_\upsilon &= \mathbb{E}_f\left[\mathcal{Q}_\upsilon\,\mathbb{I}_{\mathbb{A}} + \mathcal{Q}_\upsilon\,\mathbb{I}_{\mathcal{I}_\upsilon>2\sigma,\mathbb{A}^c} + \mathcal{Q}_\upsilon\mathbb{I}_{\mathcal{I}_\upsilon\le 2\sigma,\mathbb{A}^c}\right]\mathbb{I}_G \\ &\le \mathbb{P}_f\left(\mathbb{A},G\right) + \mathbb{P}_f\big(\mathcal{I}_\upsilon > 2\sigma\big) + 4\delta^{-D_b}\sigma.\end{aligned}$$

Under the event G, we notice that $[0,\delta]^{D_b} \subseteq nh^d\big(\Theta\big(A(f),M(f)\big)-\theta\big) \subseteq \hat{\Upsilon}_n$ for any $\delta \le \big(2M(f)-A(f)\big)$. Using Lemma 1 (*3*) and the inequality (5.16), we have

$$\mathbb{E}_f\mathcal{Q}_\upsilon \le 2\mathcal{C}_h\delta + T^{D_b}\sqrt{|\Omega_\upsilon|}\sigma^{-1/2}e^{-g_h(\upsilon)} + \mathcal{C}_h|\Omega_\upsilon|T^{-1}\sigma^{-1} + 4\delta^{-D_b}\sigma.$$

Choosing $T = \exp\Big\{\frac{1}{2D_b}g_h(\upsilon)\Big\}$, $\sigma = \exp\Big\{-\frac{1}{3D_b}g_h(\upsilon)\Big\}$ and $\delta = \exp\Big\{-\frac{1}{6D_b^2}g_h(\upsilon)\Big\}$, we obtain

$$\mathbb{E}_f\mathcal{Q}_\upsilon \le \left[2\mathcal{C}_h + |\Omega_\upsilon|\left((\upsilon+1)^{D_b} + \mathcal{C}_h\right) + 4\right]\exp\left\{-\frac{1}{6D_b^2}g_h(\upsilon)\right\}.$$

Conclusion of the proof of Assertion 2. The simplest algebra shows that $|\Omega_\upsilon| \le \big(2\upsilon+2\big)^{D_b}$, so we get

$$\mathbb{E}_f\mathcal{Q}_\upsilon \quad \le \quad \left[\upsilon^{D_b+1} + 2(2\upsilon+2)^{2D_b} + 5\right]\mathcal{C}_h\ \exp\left\{-\frac{1}{6D_b^2}g_h(\upsilon)\right\}, \tag{5.17}$$

Note that if the event G is realized then $\hat{\Upsilon}_n(a) \subseteq \Upsilon_n(a) = \bigcup_{j=0}^{\infty} \Omega_{a+j}$. We obtain in view of (5.17)

$$\begin{aligned}\mathbb{E}_f \int_{\hat{\Upsilon}_n \cap (\|u\|_1 > a)} \|u\|_1 z_h(u)\, \mathbb{I}_G\, du &\leq \sum_{j=0}^{\infty} \big(a+j+1\big)\mathbb{E}_f\, \mathcal{Q}_{a+j} \\ &= \Sigma\big(M(f)\big)\, a\, \mathcal{B}_a\, \mathcal{C}_h\, \exp\left\{-\frac{1}{6D_b^2} g_h(a)\right\}.\end{aligned}$$

where we have put $\mathcal{B}_a = a^{D_b+1} + 2(2\upsilon+2)^{2D_b} + 5 + D_b\left(a^{D_b} + (2a+2)^{\frac{D_b}{2}-1}\right)$.
∎

5.3. *Proof of Proposition 2*

To prove the proposition it suffices to integrate the inequality obtained in Proposition 1 and to use the following lemma which will be extensively exploited in the sequel.

Lemma 2. *There exists $\lambda > 0$ such that $\forall n > 1$ and $\forall h \in \mathcal{H}_n$, we have*

$$\lambda_n(h) \geq \lambda.$$

where $\lambda_n(h)$ is the smallest eigenvalue of the matrix

$$\mathcal{M}_{nh}(y) = \frac{1}{nh^d} \sum_{i=1}^{n} K^\top\left(\frac{X_i - y}{h}\right) K\left(\frac{X_i - y}{h}\right) \mathbb{I}_{V_h(y)}(X_i),$$

and $K(z)$ is the D_b-dimensional vector of the monomials z^p, $p \in \mathcal{P}_b$.

Proof of Proposition 2. In order to simplify the proof, let us introduce the following constants

$$c_1 = \frac{(1+6D_b^2)}{6A(f)D_b^2}, \quad c_2 = \frac{\lambda}{432M(f)\, D_b^3}.$$

By definition of $A(f)$, $M(f)$ and $\mathfrak{B}(.,.)$ respectively in (1.10) and (5.9), we have the following inequality $\mathfrak{B}\big(A(f), M(f)\big) \leq \mathfrak{B}(A, M)$ for any $f \in \mathbb{H}_d(\beta, L, M, A)$. By integration of the inequality of Proposition 1 and using Lemma 2, we get for any $q \geq 1$ and $f \in \mathbb{H}_d(\beta, L, M, A)$

$$\begin{aligned}
&\mathbb{E}_f\big|\hat{f}^h(y) - f(y)\big|^q \mathbb{I}_G \\
&= \int_0^{+\infty} \eta^{q-1}\mathbb{P}_f\left(\big|\hat{f}^h(y) - f(y)\big| \geq \eta, G\right) d\eta \\
&= (nh^d)^{-q} \int_0^{+\infty} \eta^{q-1}\mathbb{P}_f\left(\big|\hat{f}^h(y) - f(y)\big| \geq \frac{\eta}{nh^d}, G\right) d\eta \\
&\leq (nh^d)^{-q}\left[\int_0^{\frac{2c_1}{c_2}(1\vee\mathcal{N}_h)} \eta^{q-1}\, d\eta \right.\\
&\qquad \left. + \int_{\frac{2c_1}{c_2}(1\vee\mathcal{N}_h)}^{+\infty} \eta^{q-1}\mathbb{P}_f\left(\big|\hat{f}^h(y) - f(y)\big| \geq \frac{\eta}{nh^d}, G\right) d\eta\right] \\
&\leq \frac{(1\vee\mathcal{N}_h)^q}{(nh^d)^q}\left[\frac{2^q c_1^q}{q\, c_2^q} + \frac{2^q}{c_2^q}\mathfrak{B}\big(A(f), M(f)\big)\Gamma(q)\right],
\end{aligned}$$

where $\Gamma(\cdot)$ is the well-known Gamma function. By definition of b_h and $\mathcal{N}_h$ respectively defined in (5.3) and (5.4), the assertion of Proposition 2 is proved:

$$\mathbb{E}_f\big|\hat{f}^h(y) - f(y)\big|^q \mathbb{I}_G \leq C_q^*\big(A(f), M(f)\big)\left[\frac{1 \vee Ld\, nh^{\beta+d}}{nh^d}\right]^q,$$

where

$$C_q^*(a,m) = \frac{1}{q}\left[\frac{432mD_b^3(1+6D_b^2)}{3\lambda a D_b^2}\right]^q + \left[864m\lambda^{-1}D_b^3\right]^q \mathfrak{B}(a,m)\Gamma(q), \quad a, m > 0.$$

∎

5.4. Proof of Theorem 2.2

By definition of $\bar{h} = (Ln)^{-\frac{1}{\beta+d}}$ we have

$$Ldn\bar{h}^{\beta+d} = d, \quad (n\bar{h}^d)^{-q} = L^{\frac{qd}{\beta+d}}\varphi_n^q(\beta).$$

Applying the inequality given in Remark 4, we come to the assertion of the theorem. ∎

5.5. Proof of Theorem 3.3

This Proof is based on the Lepski scheme developed by [12] and adapted for the bandwidth selection by [15]. We start the proof with formulating auxiliary Lemmas whose proofs are given in Appendix (Section 7). Define

$$h^* = \left[n^{-1}c\left(1 + \frac{(b-\beta)}{(b+d)(\beta+d)}\ln n\right)\right]^{\frac{1}{\beta+d}},$$

where the positive constant c is chosen as follows

$$c < [1 \wedge 1/(Ld)]\,[1 \wedge 4/M(f)]\,\frac{\beta+d-1}{\beta+d}\left[1 \wedge \frac{A}{144MD_b}\right]\left[1 \wedge \frac{6AD_b^2}{1+6D_b^2}\right],$$

and let the integer κ be defined as follows.

$$2^{-\kappa}h_{\max} \leq h^* < 2^{-\kappa+1}h_{\max}. \tag{5.18}$$

The definitions of h^* and κ imply the following Lemmas.

Lemma 3.

$$\mathbb{E}_f\big|\hat{f}^{(k)}(y) - f(y)\big|^q \mathbb{I}_G \leq \bar{C}_q \frac{(1+k\ln 2)^q}{(nh_k^d)^q}, \quad \forall k \geq \kappa,$$

where $\bar{C}_q = C_q^*\big(A(f), M(f)\big)\dfrac{c(\beta+d)}{(\beta+d-1)(Ld)^{-1}}$.

Lemma 4. *For any* $f \in \mathbb{H}_d(\beta, L, M, A)$ *and any* $k \geq \kappa+1$

$$\mathbb{P}_f\big(\hat{k}=k,\ G\big) \leq J_2\mathfrak{B}(A,M)\exp\big\{J_1 n (h^*)^{\beta+d}\big\}2^{-(k-1)(8qd+4)},$$

where $J_1 = Ld(1+6D_b^2)/6AD_b^2$ *and* $J_2 = (1-2^{-(8qd+4)})^{-1}$.

Lemma 5. *There exists a universal constant* $\vartheta > 0$ *such that*

$$\limsup_{n\to\infty} \sup_{f \in \mathbb{H}_d(\beta,L,M,A)} \exp\left\{\frac{An^{\frac{b}{b+d}}}{16M\vartheta^2 D_b^2}\right\}\mathbb{P}_f\big(G^c\big) = 0.$$

Proof of Theorem 3.3. We decompose the risk as follows

$$\begin{aligned}
&\mathbb{E}_f\big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_G \\
&\quad \leq \mathbb{E}_f\big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_{\hat{k}\leq\kappa,G} + \mathbb{E}_f\big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_{\hat{k}>\kappa,G} \\
&\quad := R_1(f) + R_2(f).
\end{aligned} \tag{5.19}$$

First we control R_1. Obviously

$$\big|\hat{f}^{(\hat{k})}(y) - f(y)\big| \leq \big|\hat{f}^{(\hat{k})}(y) - \hat{f}^{(\kappa)}(y)\big| + \big|\hat{f}^{(\kappa)}(y) - f(y)\big|.$$

Note that the realization of the event G implies $\hat{M} \leq 3M(f)/2$. This together with the definition of $\hat{k}$ yields

$$\big|\hat{f}^{(\hat{k})}(y) - \hat{f}^{(\kappa)}(y)\big|\mathbb{I}_{\hat{k}\leq\kappa,G} \leq Cs_n(\kappa), \quad s_n(k) = (1+k\ln 2)^q (nh_k^d)^{-q},$$

where $C = 288MD_b^3\lambda^{-1}(32qd+16)$. In view of Lemma 3 we also get

$$\mathbb{E}_f\big|\hat{f}^{(\kappa)}(y) - f(y)\big|^q \leq \bar{C}_q s_n(\kappa).$$

Noting that the right hand side of the obtain inequality is independent of f and taking into account the definition of κ and h^* we obtain

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\beta,L,A,M)} \phi_n^{-q}(\beta) R_1(f) < \infty. \tag{5.20}$$

Now we will find a bound of R_2. Applying Cauchy-Schwartz inequality we have in view of Lemma 4

$$\begin{aligned} R_2(f) &= \sum_{k>\kappa}^{\mathrm{k}_n} \mathbb{E}_f \big|\hat{f}^{(k)}(y) - f(y)\big|^q \mathbb{I}_{\left[\hat{k}=k,G\right]} \\ &\leq \sum_{k>\kappa} \Big(\mathbb{E}_f \big|\hat{f}^{(k)}(y) - f(y)\big|^{2q}\Big)^{1/2} \sqrt{\mathbb{P}_f\{\hat{k}=k, G\}} \\ &= \Delta(h^*) \sum_{k>\kappa} \Big(\mathbb{E}_f \big|\hat{f}^{(k)}(y) - f(y)\big|^{2q}\Big)^{1/2} 2^{-(k-1)(4qd+2)}, \end{aligned} \tag{5.21}$$

where we have put $\Delta(h^*) = J_2 \mathfrak{B}(A,M) \exp\left\{J_1 n (h^*)^{\beta+d}\right\}$. We obtain from Lemma 3 and (5.21)

$$R_2(f) \leq \; J_3 \, (n h_{\max}^d)^{-q} \exp\left\{J_1 n (h^*)^{\beta+d}\right\}, \tag{5.22}$$

where

$$J_3 = J_2 \mathfrak{B}(A,M) \, 2^{4qd+2} \, \bar{C}_{2q}^{1/2} \sum_{s\geq 0} (1 + s\ln 2)^q 2^{-3sdq-2}.$$

It remains to note that the definition of h^* implies that

$$\limsup_{n\to\infty} \phi_n^{-q}(\beta) (n h_{\max}^d)^{-q} \exp\left\{J_1 n (h^*)^{\beta+d}\right\} < \infty$$

and that the right hand side of (5.22) is independent of f. Thus,we have

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\beta,L,A,M)} \phi_n^{-q}(\beta) R_2(f) < \infty.$$

that yields together with (5.19) and (5.20)

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\beta,L,A,M)} \phi_n^{-q}(\beta) \mathbb{E}_f \big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_G < \infty.$$

To get the assertion of the theorem it suffices to show that

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\beta,L,A,M)} \phi_n^{-q}(\beta) \mathbb{E}_f \big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_{G^c} < \infty. \tag{5.23}$$

Note that $\hat{f}^{(\hat{k})}(y) \leq 6\hat{M}$ in view of (3.4). Note also that the local least square estimator $\tilde{\delta}$ is a linear function of observation $Y^{(n)}$ and moreover $0 \leq Y_i \leq M, i = 1, ..., n$. This together with the definition of $\hat{M}$, (expression (3.3)) allows us to state that there exists $0 < J_4 < +\infty$ such that $\big|\hat{f}^{(\hat{k})}(y) - f(y)\big| \le J_4 M$. Here we also have taken into account that $||f||_\infty \le M$.

Finally we obtain

$$\mathbb{E}_f\big|\hat{f}^{(\hat{k})}(y) - f(y)\big|^q \mathbb{I}_{G^c} \le J_4^q\, M^q \mathbb{P}_f\big\{G^c\big\}.$$

and (5.23) follows now from Lemma 4. ∎

6. Proofs of lower bounds

The proofs of Theorems 2.1 and 3.2 are based on the following proposition.

Put $\phi_n(\gamma) = \big[n^{-1}\big(1 + (b-\gamma)\ln n\big)\big]^{\frac{\gamma}{\gamma+d}}$, $\gamma \in (0, b]$ and let

$$R_n^{(q)}(\tilde{f}, v) = \sup_{f\in\mathbb{H}_d(\alpha,L,M,A)} \mathbb{E}_f\left[\phi_n^{-q}(\alpha)\,|\tilde{f}(y) - f(y)|^q\right] + \sup_{f\in\mathbb{H}_d(\beta,L,M,A)} \mathbb{E}_f\left[n^{-vq}\phi_n^{-q}(\beta)\,|\tilde{f}(y) - f(y)|^q\right].$$

where $v \ge 0$ and $\alpha, \beta \in (0, b]^2$.

Proposition 3. *Let Ψ be admissible family of normalizations such that*

$$\psi_n(\alpha)/\phi_n(\alpha) \xrightarrow[n\to\infty]{} 0.$$

Then, for any $0 \le v < (\beta-\alpha)/(\beta+1)(\alpha+1)$

$$\liminf_{n\to\infty} \inf_{\tilde{f}} R_n^{(q)}(\tilde{f}, v) > 0.$$

The proof is given in section 6.3.

6.1. Proof of Theorem 2.1

Using the proposition 3 for $\beta = \alpha$, we have to choose $v = 0$ and one gets

$$\begin{aligned} R_{n,q}\big[\mathbb{H}_d(\beta, L, M, A)\big] &= R_n^{(q)}(\tilde{f}, 0) \\ &= \sup_{f\in\mathbb{H}_d(\alpha,L,M,A)} \mathbb{E}_f\left[n^{-q\frac{\alpha}{\alpha+1}}\big|\tilde{f}(y) - f(y)\big|^q\right] > 0, \quad \forall \tilde{f}. \end{aligned}$$

∎

6.2. Proof of Theorem 3.2

I. To prove the first assertion of the theorem it suffices to consider the family $\{\upsilon_n(\beta)\}_{\beta\in(0,b]}$, where $\upsilon_n(\alpha) = \varphi_n(\alpha)$ and $\upsilon_n(\beta) = 1$ for any $\beta \ne \alpha$. The corresponding attainable estimator is the estimator being minimax on $\mathbb{H}_d(\alpha, L, M, A)$.

II. Let us consider the family $\{\phi_n(\beta)\}_{\beta\in]0,b]}$, which is admissible in view of Theorem 3.3. First, we note that $\gamma = b$ is not possible since $\phi_n(b) = \varphi_n(b)$ is the minimax rate of convergence on $\mathbb{H}_d(b, L, M, A)$.

Thus we assume that $\gamma \in]0, b[$ satisfying (3.2). Let $\hat{f}^\Psi$ be a $\Psi^{(n)}$-attainable estimator. Since $\psi_n(\gamma)/\phi_n(\gamma) \to 0, n \to \infty$ in view of (3.2) then obviously

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\gamma,L,M,A)} \mathbb{E}_f\left[\phi_n^{-q}(\gamma)\left|\hat{f}^\Psi(y) - f(y)\right|^q\right] = 0.$$

Therefore, applying Proposition 3 with $v = 0$ we have for any $\beta < \gamma$

$$\limsup_{n\to\infty} \sup_{f\in\mathbb{H}_d(\beta,L,M,A)} \mathbb{E}_f\left[\phi_n^{-q}(\beta)\left|\hat{f}^\Psi(y) - f(y)\right|^q\right] > 0.$$

We conclude that necessarily $\psi_n(\beta) \gtrsim \phi_n(\beta)$ for any $\beta < \gamma$.

Moreover for any $\beta > \gamma$, applying Proposition 3 with an arbitrary $0 \le v < (\beta-\gamma)/(\beta+1)(\gamma+1)$ we obtain that

$$\psi_n(\beta) \gtrsim n^v \phi_n(\beta), \ \ \beta > \gamma.$$

It remains to note that the form of rate of convergence proved in Theorem 2.1 implies that

$$\phi_n(\gamma)/\psi_n(\gamma) = o\left([\ln n]^{\frac{\gamma}{\gamma+d}}\right).$$

∎

6.3. Proof of Proposition 3

Let $\varkappa > 0$ a parameter whose choice will be done later. Put

$$h = \left(\varkappa \frac{1 + (\beta-\alpha)\ln n}{n}\right)^{\frac{1}{\alpha+d}}.$$

Without loss of generality we will assume later that $L > 1$.

Consider the functions: $f_0 \equiv 1$ and

$$f_1(x) = 1 - \big(L-1\big)\varkappa^{\frac{\alpha}{\alpha+d}}\phi_n(\alpha)F\left(\frac{x_1-y_1}{h}, ..., \frac{x_d-y_d}{h}\right), \quad x \in [0,1]^d.$$

Here F is a compactly supported positive function belonging to $\mathbb{H}_d(\alpha, 1, M, A)$ such that $F(0) = 1 = \max_x F(x)$.

We can easily see that $f_1 \in \mathbb{H}_d(\alpha, L, M, A)$. Therefore, we have

$$\begin{aligned} R_n^{(q)}(\tilde{f}, v) \quad &\ge \quad \mathbb{E}_0\left|n^{-v}\phi_n^{-1}(\beta)\big(\tilde{f}(y)-1\big)\right|^q + \mathbb{E}_1\left|\phi_n^{-1}(\alpha)\big(\tilde{f}(y)-f_1(y)\big)\right|^q \\ &\ge \quad \mathbb{E}_0\left|n^{-v}\phi_n^{-1}(\beta)\big(\tilde{f}(y)-1\big)\right|^q + \mathbb{E}_1\left|\phi_n^{-1}(\alpha)\big(\tilde{f}(y)-1\big)+z\right|^q, \end{aligned}$$

where $z = \big(L-1\big)\varkappa^{\frac{\alpha}{\alpha+1}}F(0)$. Set

$$\tilde{\lambda} = \phi_n^{-1}(\alpha)\big(1-\tilde{f}(y)\big), \quad \varsigma_n = n^{-v}\frac{\phi_n(\alpha)}{\phi_n(\beta)} = n^{-v}\left(\frac{\ln n}{n}\right)^{-\varrho},$$

where $\varrho = \frac{\beta-\alpha}{(\beta+1)(\alpha+1)}$. We get

$$\begin{aligned} R_n^{(q)}(\tilde{f},v) &\geq \mathbb{E}_0\big|\varsigma_n\tilde{\lambda}\big|^q + \mathbb{E}_1\big|z-\tilde{\lambda}\big|^q \\ &\geq \mathbb{E}_0\big|\varsigma_n\tilde{\lambda}\big|^q\mathbb{I}_{\{|\tilde{\lambda}|>z/2\}} + \mathbb{E}_1\big|z-\tilde{\lambda}\big|^q\mathbb{I}_{\{|\tilde{\lambda}|\leq z/2\}} \\ &\geq \mathbb{E}_0\big|\varsigma_n\frac{z}{2}\big|^q\mathbb{I}_{\{|\tilde{\lambda}|>z/2\}} + \mathbb{E}_1\big|\frac{z}{2}\big|^q\mathbb{I}_{\{|\tilde{\lambda}|\leq z/2\}}. \end{aligned}$$

Noting that $f_1 \leq f_0$, since F is positive, and putting $c_n\big(Y^{(n)}\big) = \mathbb{I}_{\{|\tilde{\lambda}|>z/2\}}$ we obtain

$$\begin{aligned} R_n^{(q)}(\tilde{f},v) \geq \varsigma_n^q\frac{z^q}{2^q}\frac{\prod_{i=1}^n f_1(X_i)}{\prod_{i=1}^n f_1(X_i)}\int_0^{f_1(X_1)}\dots\int_0^{f_1(X_n)} c_n(x)dx_1\dots dx_n \\ +\frac{z^q}{2^q}\frac{1}{\prod_{i=1}^n f_1(X_i)}\int_0^{f_1(X_1)}\dots\int_0^{f_1(X_n)} 1-c_n(x)dx_1\dots dx_n. \quad (6.1) \end{aligned}$$

We have

$$\begin{aligned} \prod_{i=1}^n f_1(X_i) &= \prod_{i=1}^n\left(1-(L-1)\varkappa^{\frac{\alpha}{\alpha+d}}\phi_n(\alpha)F\left(\frac{X_i-y}{h}\right)\right) \\ &\geq \big(1-(L-1)\varkappa^{\frac{\alpha}{\alpha+d}}\phi_n(\alpha)\big)^{nh^d} \geq e^{-(L-1)\varkappa}n^{-(L-1)\varkappa(\beta-\alpha)}. \quad (6.2) \end{aligned}$$

We obtain in view of (6.1) and (6.2)

$$\begin{aligned} R_n^{(q)}(\tilde{f},v) &\geq \varsigma_n^q\frac{z^q}{2^q}e^{-(L-1)\varkappa}n^{-(L-1)\varkappa(\beta-\alpha)} \\ &\quad\times\frac{1}{\prod_{i=1}^n f_1(X_i)}\int_0^{f_1(X_1)}\dots\int_0^{f_1(X_n)} c_n(x)dx_1\dots dx_n \\ &\quad+\frac{z^q}{2^q}\frac{1}{\prod_{i=1}^n f_1(X_i)}\int_0^{f_1(X_1)}\dots\int_0^{f_1(X_n)} 1-c_n(x)dx_1\dots dx_n \\ &\geq \frac{z^q}{2^q}\left(1\wedge\varsigma_n^q e^{-(L-1)\varkappa}n^{-(L-1)\varkappa(\beta-\alpha)}\right). \end{aligned}$$

<u>*Case 1:*</u> $\beta=\alpha$. Choosing $\varkappa=1$, and noting that $\varsigma_n=1$ and $\prod_{i=1}^n f_1(X_i) \geq e^{-(L-1)}$, we deduce from (6.1) that

$$\inf_{\tilde{f}} R_n^{(q)}(\tilde{f},v) \geq \frac{(L-1)^q}{2^q}e^{-(L-1)} > 0.$$

<u>*Case 2:*</u> $\beta>\alpha$. Put

$$\varkappa = \frac{q\big(\varrho-v\big)-t_n}{1+(L-1)(\beta-\alpha)} > 0, \;\; t_n = \frac{q}{\ln n}\ln\frac{1}{\big(1+(\beta-\alpha)\ln n\big)^{-\varrho}} \xrightarrow[n\to\infty]{} 0.$$

This choice provides us with the following bound

$$\begin{aligned}\varsigma_n^q e^{-(L-1)\varkappa} n^{-(L-1)\varkappa(\beta-\alpha)} &= \big(1+(\beta-\alpha)\ln n\big)^{-q\varrho} e^{-(L-1)\varkappa} n^{q(\varrho-v)-(L-1)\varkappa(\beta-\alpha)} \\ &\geq \big(1+(\beta-\alpha)\ln n\big)^{-q\varrho} e^{-\frac{2}{3}q(L-1)} n^{t_n} \geq e^{-\frac{2}{3}q(L-1)}.\end{aligned}$$

This yields

$$\inf_{\tilde f} R_n^{(q)}(\tilde f, v) \geq \frac{(L-1)^q \varkappa^{q\frac{\alpha}{\alpha+1}}}{2^q} e^{-\frac{2}{3}q(L-1)} > 0.$$

∎

7. Appendix

7.1. Proof of Lemma 1

Without loss generality we will suppose later that $nh^d \in \mathbb{N}^*$. In order to simplify understanding of this proof, we note the approximation polynomial $\mathcal{A}_u^i = f_{\theta+u(nh^d)^{-1}}(X_i),\ i=1,\dots,n$ for all $u \in \Upsilon_n$.

1. Note that for $u \in \Upsilon_n$

$$\mathbb{E}_f Z_{h,\theta}(u) \leq \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_0^i}{f(X_i)} \leq e^{\mathcal{N}_h/A(f)}. \tag{7.1}$$

The first inequality is the consequence of the definition of $Z_{h,\theta}$ in (5.8) and the following calculation

$$\mathbb{E}_f \mathbb{I}_{[Y_i \leq\ \mathcal{A}_u^i]} = \mathbb{P}_f\left(Y_i \leq\ \mathcal{A}_u^i\right) = 1 \wedge \frac{\mathcal{A}_u^i}{f(X_i)}.$$

In (7.1), the second inequality is obtained with classical inequality $1+\rho \leq e^\rho$, $\rho \in \mathbb{R}$ and having on mind that $\forall x \in V_h(y),\ f_\theta(x) \geq f(x)$,

$$\prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_0^i}{f(X_i)} = \prod_{i:\ X_i \in V_h(y)} \left(1 + \frac{\mathcal{A}_0^i - f(X_i)}{f(X_i)}\right) \leq \exp\left\{\frac{b_h \times nh^d}{A(f)}\right\}$$

Case 1: If $\|u_1-u_2\|_1 \geq 1$, the inequality (7.1) allows to get

$$\mathbb{E}_f\big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big| \leq \mathbb{E}_f Z_{h,\theta}(u_1) + \mathbb{E}_f Z_{h,\theta}(u_2) \leq 2e^{\mathcal{N}_h/A(f)}\|u_1-u_2\|_1.$$

Case 2: Assume now that $\|u_1-u_2\|_1 < 1$ and introduce the random events

$$\begin{aligned}F_1 &= \left\{\forall i=1,\dots,n:\ Y_i \leq \mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i\right\}, \\ F_2 &= \left\{\forall i=1,\dots,n:\ Y_i \leq \mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i\right\} \\ &\quad \cap \left\{\exists i:\ Y_i > \mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i\right\}, \\ F_3 &= \left\{\exists i:\ Y_i > \mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i\right\}.\end{aligned}$$

We have used the following notations: $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$, $a,b \in \mathbb{R}$. For any $(u_1,u_2) \in \Upsilon_n^2$, we have

$$\begin{aligned}
&\mathbb{E}_f\big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big| = \mathbb{E}_f\big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big|\,\mathbb{I}_{[F_1]}\\
&+\mathbb{E}_f\big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big|\,\mathbb{I}_{[F_2]} + \mathbb{E}_f\big|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\big|\,\mathbb{I}_{[F_3]}\\
&= \mathcal{K}_1 + \mathcal{K}_2 + \mathcal{K}_3. \qquad (7.2)
\end{aligned}$$

The following bound will be extensively exploited in the sequel.

$$A(f)/4 \le f_v(x) \le 9M(f), \quad \forall v \in \Theta\big(A(f)/4, 3M(f)\big),\ x \in V_h(y).$$

Control of $\mathcal{K}_1$.

$$\mathcal{K}_1 = \left|\prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_0^i}{\mathcal{A}_{u_1}^i} - \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_0^i}{\mathcal{A}_{u_2}^i}\right| \mathbb{P}_f\{F_1\}, \qquad (7.3)$$

and

$$\mathbb{P}_f\{F_1\} = \prod_{i:\ X_i \in V_h(y)} \mathbb{P}_f\{Y_i \le \mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i\} \le \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i}{f(X_i)}. \qquad (7.4)$$

Therefore, using (7.1), we have

$$\begin{aligned}
\mathcal{K}_1 \quad &\le \quad \left(1 - \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i}{\mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i}\right) \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}_0^i}{f(X_i)}\\
&\le \quad e^{\mathcal{N}_h/A(f)} \left(1 - \exp\left\{\sum_{i:\ X_i \in V_h(y)} \ln \frac{\mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i}{\mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i}\right\}\right). \qquad (7.5)
\end{aligned}$$

Remember that $\left|\mathcal{A}_{u_1}^i - \mathcal{A}_{u_2}^i\right| \le (nh^d)^{-1}\|u_1 - u_2\|_1$ and $\mathcal{A}_u^i \ge A(f)/4$. Let us give the following calculation with inequality of finite increments for $\ln(\cdot)$

$$\ln \frac{\mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i}{\mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i} = -\left|\ln \mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i - \ln \mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i\right| \ge -\frac{(nh^d)^{-1}\|u_1 - u_2\|_1}{\mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i}$$

Using last inequalities, (7.3), (7.4), (7.5), last inequality and the well known inequality $1 - e^{-\rho} \le \rho$, we have

$$\mathcal{K}_1 \le \frac{4}{A(f)} e^{\mathcal{N}_h/A(f)} \|u_1 - u_2\|_1.$$

Control of $\mathcal{K}_2$. F_2 can be rewritten as

$$\begin{aligned}
F_2 \quad &= \quad \big\{\forall i = 1,\dots,n:\ Y_i \le \mathcal{A}_{u_1}^i \vee \mathcal{A}_{u_2}^i\big\}\\
&\qquad \setminus \big\{\forall i = 1,\dots,n:\ Y_i \le \mathcal{A}_{u_1}^i \wedge \mathcal{A}_{u_2}^i\big\}\\
&= \quad G \setminus F_1.
\end{aligned}$$

and define

$$\begin{aligned}\mathcal{G}_1 &= \left\{X_i \in V_h(y):\ \mathcal{A}^i_{u_1} \vee \mathcal{A}^i_{u_2} < f(X_i)\right\},\\ \mathcal{G}_2 &= \left\{X_i \in V_h(y):\ \mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2} < f(X_i)\right\}.\end{aligned}$$

Note that $F_1 \subseteq G$ and, therefore,

$$\begin{aligned}\mathcal{K}_2 &\leq \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}^i_0}{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}} \left(\mathbb{P}_f\{G\} - \mathbb{P}_f\{F_1\}\right)\\ &= \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}^i_0}{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}} \left(\prod_{i:\ X_i \in \mathcal{G}_1} \frac{\mathcal{A}^i_{u_1} \vee \mathcal{A}^i_{u_2}}{f(X_i)} - \prod_{i:\ X_i \in \mathcal{G}_2} \frac{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}}{f(X_i)}\right)\end{aligned}$$

The definition of $\mathcal{G}_2$ implies

$$\prod_{i:\ X_i \in V_h(y)} \frac{1}{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}} \leq \prod_{i:\ X_i \in \mathcal{G}_2} \frac{1}{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}} \prod_{i:\ X_i \in \mathcal{G}_2^c} \frac{1}{f(X_i)}$$

Since $\mathcal{G}_1 \subseteq \mathcal{G}_2$, $\|u_1 - u_2\|_1 < 1$ and $|f_u(x)| \leq ||u||_1$, $\forall x \in [0,1]^d$, $\forall u \in \Upsilon_n$, using the last inequality and (7.1), we obtain

$$\begin{aligned}\mathcal{K}_2 &\leq \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}^i_0}{f(X_i)} \left(\prod_{i:\ X_i \in \mathcal{G}_2} \frac{\mathcal{A}^i_{u_1} \vee \mathcal{A}^i_{u_2}}{\mathcal{A}^i_{u_1} \wedge \mathcal{A}^i_{u_2}} - 1\right)\\ &\leq 4D_b e^{1+\mathcal{N}_h/A(f)} \|u_1 - u_2\|_1 / A(f).\end{aligned}$$

Control of $\mathcal{K}_3$. We can rewrite the process $Z_{h,\theta}$ with the notation $\mathcal{A}^i_u$

$$Z_{h,\theta}(u) = \prod_{i:\ X_i \in V_h(y)} \frac{\mathcal{A}^i_0}{\mathcal{A}^i_u} \mathbb{I}_{[Y_i \leq \mathcal{A}^i_u]}.$$

Under the event F_3, we get

$$\left|Z_{h,\theta}(u_1) - Z_{h,\theta}(u_2)\right| \mathbb{I}_{[F_3]} = 0$$

Then $\mathcal{K}_3 = 0$.

The first assertion of the lemma is proved with (7.2) and the bounds of $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$.

2. For any $u \in \Upsilon_n$, since the random variables $(Y_i)_i$ are independent we have,

$$\mathbb{E}_f Z^{1/2}_{h,\theta}(u) = \prod_{i:\ X_i \in V_h(y)} \sqrt{\frac{\mathcal{A}^i_0}{\mathcal{A}^i_u}}\, \mathbb{P}_f\left\{Y_i \leq \mathcal{A}^i_u\right\}.$$

For any i, we have

$$\sqrt{\frac{\mathcal{A}_0^i}{\mathcal{A}_u^i}}\,\mathbb{P}_f\left\{Y_i\leq\mathcal{A}_u^i\right\}=\sqrt{\frac{\mathcal{A}_0^i}{\mathcal{A}_u^i}}\left[1\wedge\frac{\mathcal{A}_u^i}{f(X_i)}\right]\leq\frac{\mathcal{A}_0^i}{f(X_i)}\left[\frac{f(X_i)}{\sqrt{\mathcal{A}_0^i}\sqrt{\mathcal{A}_u^i}}\wedge\frac{\sqrt{\mathcal{A}_u^i}}{\sqrt{\mathcal{A}_0^i}}\right].$$

In view of (5.6) and (5.7), recall that $A(f)\leq f(x)\leq f_\theta(x)\leq 3M(f),\forall x\in V_h(y)$. Moreover, for any $u\in\Upsilon_n=nh^d\big(\Theta\big(A(f)/4,3M(f)\big)-\theta\big)$,

$$A(f)/4\leq f_{\theta+u(nh^d)^{-1}}(x)\leq 9M(f),\quad\forall x\in V_h(y).$$

Thus for all $i\,:\,X_i\in V_h(y)$,

$$\sqrt{\frac{\mathcal{A}_0^i}{\mathcal{A}_u^i}}\,\mathbb{P}_f\left\{Y_i\leq\mathcal{A}_u^i\right\}\leq\frac{\mathcal{A}_0^i}{f(X_i)}\left[\frac{\sqrt{\mathcal{A}_0^i}}{\sqrt{\mathcal{A}_u^i}}\wedge\frac{\sqrt{\mathcal{A}_u^i}}{\sqrt{\mathcal{A}_0^i}}\right]\leq\frac{\mathcal{A}_0^i}{f(X_i)}\left[1-\frac{\left|\mathcal{A}_0^i-\mathcal{A}_u^i\right|}{9M(f)}\right]^{1/2}.$$

The last inequality implies

$$\begin{aligned}\mathbb{E}_f Z_{h,\theta}^{1/2}(u) &\leq \prod_{i:\,X_i\in V_h(y)}\frac{\mathcal{A}_0^i}{f(X_i)}\sqrt{1-\frac{|f_{u(nh^d)^{-1}}(X_i)|}{9M(f)}}\\ &\leq e^{\mathcal{N}_h/A(f)}\exp\left\{-\frac{1}{18M(f)\,nh^d}\sum_{i:\,X_i\in V_h(y)}\left|f_u(X_i)\right|\right\}. \end{aligned}\tag{7.6}$$

It remains to show

$$\frac{1}{nh^d}\sum_{i:\,X_i\in V_h(y)}\left|f_u(X_i)\right|\geq\lambda_n(h)D_b^{-1}||u||_1.\tag{7.7}$$

Let us remember that $u=(u_p,\,p\in\mathcal{P}_b)$ (where $\mathcal{P}_b$ is defined in (1.4)). First, we get from the definition of f_u

$$f_u(x)=u\,K^\top\left(\frac{x-y}{h}\right)=K\left(\frac{x-y}{h}\right)\,u^\top,\quad\forall x\in[0,1]^d,$$

and therefore,

$$\frac{1}{nh^d}\sum_{i:\,X_i\in V_h(y)}\left|f_u(X_i)\right|=\frac{1}{nh^d}\sum_{i:\,X_i\in V_h(y)}\left|u\,K^\top\left(\frac{X_i-y}{h}\right)\right|.$$

Assume $u\neq0$ and put $v=u/||u||_1$. Noting that $|f_v(x)|\leq1,\quad\forall x\in[0,1]^d$, we have

$$
\begin{aligned}
&\frac{1}{nh^d}\sum_{i:\, X_i\in V_h(y)}\big|f_u(X_i)\big|\\
&\geq \frac{1}{nh^d}\sum_{i:\, X_i\in V_h(y)}\left|u\,K^\top\left(\frac{X_i-y}{h}\right)\right|\,|f_v(X_i)|\\
&= \frac{1}{||u||_1 nh^d}\sum_{i:\, X_i\in V_h(y)}\left|u\,K^\top\left(\frac{X_i-y}{h}\right)K\left(\frac{X_i-y}{h}\right)\,u^\top\right|\\
&\geq \frac{1}{||u||_1}\left|u\,\frac{1}{nh^d}\sum_{i:\, X_i\in V_h(y)}K^\top\left(\frac{X_i-y}{h}\right)K\left(\frac{X_i-y}{h}\right)\,u^\top\right|.
\end{aligned}
$$

The bound (7.7) follows now from Lemma 2. The assertion of the lemma follows from (7.6) and (7.7).

3. In view of Lemma 1 (*1*), we have

$$
\mathbb{E}_f\big|Z_{h,\theta}(u)-Z_{h,\theta}(0)\big|\leq \mathcal{C}_h\|u\|_1,\quad u\in\Upsilon_n\backslash 0. \tag{7.8}
$$

Taking into account that $Z_{h,\theta}(0)=1$ we obtain applying (7.8), Fubini's theorem and Tchebychev inequality

$$
\begin{aligned}
&\mathbb{P}_f\left\{\int_0^\delta\cdots\int_0^\delta Z_{h,\theta}(v)dv<\frac{1}{2}\delta^{D_b}\right\}\\
&\quad = \mathbb{P}_f\left\{\int_0^\delta\cdots\int_0^\delta \big(Z_{h,\theta}(v)-Z_{h,\theta}(0)\big)dv<-\frac{1}{2}\delta^{D_b}\right\}\\
&\quad \leq \mathbb{P}_f\left\{\int_0^\delta\cdots\int_0^\delta \big|Z_{h,\theta}(v)-Z_{h,\theta}(0)\big|dv>\frac{1}{2}\delta^{D_b}\right\}\\
&\quad \leq 2\delta^{-D_b}\int_0^\delta\cdots\int_0^\delta \mathbb{E}_f\big|Z_{h,\theta}(v)-Z_{h,\theta}(0)\big|dv\\
&\quad \leq 2\mathcal{C}_h\delta
\end{aligned}
$$

■

7.2. Proof of Lemma 2

First step: $\mathcal{M}_{nh}(y)$ is a nonnegative positive matrix.
Let $\mathcal{H}_n, n>1$ as defined in (5.1). First, we prove that

$$
\inf_{h\in\mathcal{H}_n}\lambda_n(h)>0,\quad \forall n>1. \tag{7.9}
$$

Suppose that $\exists n_1 > 1,\ h_{n_1} \in \mathcal{H}_{n_1}$ such that $\lambda_{n_1}\big(h_{n_1}\big) = 0$. Recall that $f_t(x) = t\,K(h^{-1}(x-y))$ for all $t \in \mathbb{R}^{D_b}$ and note that $\forall \tau \in \mathbb{R}^{D_b}$

$$\begin{aligned}\tau^\top \mathcal{M}_{n_1 h_{n_1}}(y)\,\tau &= \frac{1}{nh_{n_1}^d} \sum_{i:\, X_i \in V_{h_{n_1}}(y)} \left[\tau\, K^\top\left(\frac{X_i - y}{h_{n_1}}\right)\right]^2 \\ &= \frac{1}{nh_{n_1}^d} \sum_{i:\, X_i \in V_{h_{n_1}}(y)} \big[f_\tau(X_i)\big]^2 \geq 0.\end{aligned}$$

Since $\lambda_{n_1}\big(h_{n_1}\big)$ is the smallest eigenvalue of the matrix $\mathcal{M}_{n_1 h_{n_1}}(y)$ the assumption $\lambda_{n_1}\big(h_{n_1}\big) = 0$ implies that there exist τ^* belonging to the unit sphere of $\mathbb{R}^{D_b}$ such that

$$\frac{1}{nh_{n_1}^d} \sum_{i:\, X_i \in V_{h_{n_1}}(y)} \big[f_{\tau^*}(X_i)\big]^2 = 0.$$

It obviously implies that $f_{\tau^*}(X_i) = 0$ for all $X_i \in V_{h_{n_1}}(y)$. It remains to note that $nh_{n_1}^d \geq \big(b+1\big)^d$ since $h_{n_1} \in \mathcal{H}_n$ and to apply the result obtained in [16] (page 20). It yields $\tau^* = 0$ and the obtained contradiction proves (7.9).

Second step: $\mathcal{M}_{nh}(y) \xrightarrow{n\to\infty} \mathcal{M}$.
Let λ_0 be the smallest eigenvalue of the matrix

$$\mathcal{M} = \int_{[-1/2,1/2]^d} K^\top(x)\, K(x)\, dx$$

whose general term is given by

$$\mathcal{M}_{p,q} = \prod_{j=1}^{d} \int_{-\frac{1}{2}}^{\frac{1}{2}} x_j^{p_j+q_j}\, dx_j, \quad 0 \leq |p|, |q| \leq b.$$

Let us prove that

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n} \big|\lambda_n(h) - \lambda_0\big| = 0. \tag{7.10}$$

Put $m = n^{1/d}$ and without loss of generality we will assume that m is integer. Remind that the general term of the matrix $\mathcal{M}_{nh}(y)$ is given by

$$\big(\mathcal{M}_{nh}(y)\big)_{p,q} = \frac{1}{nh^d} \sum_{i:\, X_i \in V_h(y)} \prod_{j=1}^{d} \left(\frac{X_{i_j} - y_j}{h}\right)^{p_j+q_j}.$$

where $X_{i_j} = i_j/m$ for all $j = 1, \dots, d$ and $X_i = \big(X_{i_1}, ..., X_{i_d}\big)$. We get

$$\begin{aligned}
&\frac{1}{nh^d} \sum_{i:\, X_i \in V_h(y)} \prod_{j=1}^{d} \int_{i_j-1}^{i_j} \left(\frac{x_j/m - y_j}{h}\right)^{p_j+q_j} dx_j \\
&\quad \le \frac{1}{nh^d} \sum_{i:\, X_i \in V_h(y)} \prod_{j=1}^{d} \left(\frac{X_{i_j} - y_j}{h}\right)^{p_j+q_j} \\
&\quad \le \frac{1}{nh^d} \sum_{i:\, X_i \in V_h(y)} \prod_{j=1}^{d} \int_{i_j}^{i_j+1} \left(\frac{x_j/m - y_j}{h}\right)^{p_j+q_j} dx_j,
\end{aligned}$$

It yields by change of variables that

$$\begin{aligned}
\prod_{j=1}^{d} \int_{-\frac{1}{2}-2(nh^d)^{-1}}^{\frac{1}{2}} x_j^{p_j+q_j} dx_j \quad &\le \quad \frac{1}{nh^d} \sum_{i:\, X_i \in V_h(y)} \prod_{j=1}^{d} \left(\frac{X_{i_j} - y_j}{h}\right)^{p_j+q_j} \\
&\le \quad \prod_{j=1}^{d} \int_{-\frac{1}{2}}^{\frac{1}{2}+2(nh^d)^{-1}} x_j^{p_j+q_j} dx_j, \qquad (7.11)
\end{aligned}$$

Note that $nh^d \ge \ln^{\frac{1}{1+d}}(n)$ for any $h \in \mathcal{H}_n$. This together with (7.11) yields

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_n} \left| \big(\mathcal{M}_{nh}(y)\big)_{p,q} - \mathcal{M}_{p,q} \right| = 0, \quad 0 \le |p|, |q| \le b.$$

The last result obviously imply (7.10).

Third step: *Conclusion.*
First we show that $\lambda_0 > 0$. Indeed, $\forall \tau \in \mathbb{R}^{D_b}$

$$\tau^\top \mathcal{M} \tau = \int_{[-1/2,1/2]^d} \big[f_\tau(x)\big]^2 dx \ge 0.$$

Since λ_0 is the smallest eigenvalue of the matrix $\mathcal{M}$ the assumption $\lambda_0 = 0$ would imply that there exists τ^* belonging to the unit sphere of $\mathbb{R}^{D_b}$ such that $f_{\tau^*} \equiv 0$. Since f_{τ^*} is a polynomial the last identity is possible if and only if $\tau^* = 0$. The obtained contradiction shows that $\lambda_0 > 0$.

Next, note that in view of (7.10) there exists n_0 such that $\forall n > n_0$ and $\forall h \in \mathcal{H}_n$, $\lambda_n(h) \ge \lambda_0/2$.

On the other hand in view of (7.9) $\min_{n\le n_0} \inf_{h\in\mathcal{H}_n} \lambda_n(h) > 0$. It remains to define $\lambda > 0$ as

$$\lambda = \min\left(\min_{n\le n_0} \inf_{h\in\mathcal{H}_n} \lambda_n(h),\ \lambda_0/2\right).$$

∎

7.3. *Proof of Lemma 3*

Remind that $h_k \le h_\kappa \le h^*$ by definition of h_k, h^* and κ (cf. (5.18)). Using Proposition 2 with $h = h_k$, it yields

$$\begin{aligned}
\mathbb{E}_f\big|\hat{f}^{(k)}(y) - f(y)\big|^q \mathbb{I}_G &\le C_q^*\big(A(f), M(f)\big)\left(\frac{1 \vee Ld\, nh_k^{\beta+d}}{nh_k^d}\right)^q \\
&\le C_q^*\big(A(f), M(f)\big)\left(\frac{1 \vee Ld\, n(h^*)^{\beta+d}}{nh_k^d}\right)^q . \quad (7.12)
\end{aligned}$$

The control of $n(h^*)^{\beta+d}$ requires the following calculation.

$$n(h^*)^{\beta+d} \le 1 + \frac{b-\beta}{(b+d)(\beta+d)}\ln n = \rho_n(\beta) \tag{7.13}$$

where $\rho_n(\beta)$ is the price to pay for adaptation defined in (1.9). By definition of h_k, we have

$$\begin{aligned}
1 + \kappa\ln 2 &= 1 + \ln\frac{h_{\max}}{h_k} \ge 1 + \ln\frac{h_{\max}}{h*} \\
&\ge 1 + \frac{b-\beta}{(b+d)(\beta+d)}\ln n - \frac{1}{\beta+d}\ln\big[c\big(1 + (b-\beta)\ln n\big)\big].
\end{aligned}$$

Using the classical inequality $\ln(1+x) \le x$ and $c \le 1$, we obtain with the last inequality

$$\frac{\beta+d-1}{\beta+d}\rho_n(\beta) \le 1 + \kappa\ln 2 \le 1 + k\ln 2, \forall k \ge \kappa. \tag{7.14}$$

According to (7.12), (7.13) and (7.14), Lemma 3 is proved. ∎

7.4. *Proof of Lemma 4*

Note that for any $k \ge \kappa + 1$ and by definition of $\hat{k}$ in (3.8)

$$\big\{\hat{k} = k\big\} = \cup_{l \ge k}\left\{\big|\hat{f}^{(k-1)}(y) - \hat{f}^{(l)}(y)\big| > \hat{M}S_n(l)\right\}.$$

Note that $S_n(l)$ is monotonically increasing in l and, therefore,

$$\begin{aligned}
\big\{\hat{k} = k\big\} &\subseteq \left\{\big|\hat{f}^{(k-1)}(y) - f(y)\big| > 2^{-1}\hat{M}S_n(k-1)\right\} \\
&\quad \cup \left[\cup_{l \ge k}\left\{\big|\hat{f}^{(l)}(y) - f(y)\big| > 2^{-1}\hat{M}S_n(l)\right\}\right].
\end{aligned}$$

Taking into account that the event G implies the realization of the event $\hat{M} \ge M(f)/2 \ge A/2$ we come to the following inequality: for any $k \ge \kappa + 1$

$$\begin{aligned}
\mathbb{P}\big(\hat{k} = k,\, G\big) &\le \mathbb{P}\left\{\big|\hat{f}^{(k-1)}(y) - \hat{f}(y)\big| > 4^{-1}M(f)\, S_n(k-1),\, G\right\} \\
&\quad + \sum_{l \ge k}\mathbb{P}\left\{\big|\hat{f}^{(l)}(y) - f(y)\big| > 4^{-1}M(f)\, S_n(l),\, G\right\}. \quad (7.15)
\end{aligned}$$

Now we are going to justify the use of Proposition 1. Note that $b_{h_l} \leq Ldh_l^\beta$ since $f \in \mathbb{H}_d(\beta, L, A, M)$ and, therefore, by definition of h^*, we have

$$\mathcal{N}_{h_l} \leq Ldn(h_l)^{\beta+d} \leq Ldn(h_\kappa)^{\beta+d} \leq Ldn(h^*)^{\beta+d} \leq c\rho_n(\beta), \;\; \forall l \geq k-1. \quad (7.16)$$

Remark that the definition of $S_n(l)$ yields

$$nh_l^d \, S_n(l) \geq 432 D_b^3 (32qd+16)\lambda^{-1}(h_l)\big[1+\ln\big(h_{\max}/h_l\big)\big].$$

Using (7.14), (7.16) and the last inequality, we have

$$\frac{M(f)}{4} nh_l^d \, S_n(l) \geq 144 M D_b (1 \vee \mathcal{N}_{h_l})/(\lambda_n(h_l)A). \quad (7.17)$$

The last inequality allows us to apply Proposition 1 and Lemma 2 with $\varepsilon = \frac{M(f)}{4} nh_l^d \, S_n(l)$, and we obtain $\forall l \geq k-1$

$$\begin{aligned}
&\mathbb{P}\left\{ \big|\hat{f}^{(l)}(y) - f(y)\big| > (M(f)/4)\, S_n(l), \, G \right\} \\
&\quad \leq \mathfrak{B}(A,M)\mathcal{E}(h_l)\, [h_{\max}/h_l]^{-8qd-4} \\
&\quad = \mathfrak{B}(A,M)\mathcal{E}(h_l) 2^{-l(8qd+4)}. \quad (7.18)
\end{aligned}$$

Here we have also used that $k \geq \kappa+1$. We obtain from (7.15), (7.18) and (7.16) that $k \geq \kappa+1$

$$\mathbb{P}\big(\hat{k} = k, \, G\big) \leq J_2 \mathfrak{B}(A,M) \exp\big\{J_1 n (h^*)^{\beta+d}\big\} 2^{-(k-1)(8qd+4)},$$

where $J_2 = (1-2^{-(8qd+4)})^{-1}$. ∎

7.5. Proof of Lemma 5

Set for any $p \in \mathcal{P}_b$

$$W_{ni}^p(y) = p_1!...p_d! \, \frac{h_{\max}^{d-|p|}}{n} \, K^\top(0) \, \mathcal{M}_{nh_{\max}}^{-1}(y) \, K\left(\frac{X_i - y}{h_{\max}}\right) \mathbb{I}_{V_{\max}(y)}(X_i),$$

and note that $\tilde{\delta}_p = \sum_{i=1}^d 2Y_i \, W_{ni}^p(y)$.

The model (1.1) can be rewritten as $2Y_i = f(X_i) + f(X_i)(2U_i - 1)$. Thus, setting $F(X) = \big(f(X_i)\big)_{i=1,\dots,n}$, $V(X) = \big(f(X_i)(2U_i-1)\big)_{i=1,\dots,n}$ and

$$\mathcal{D}(f) = \left(\frac{\partial^{|p|} f(y)}{\partial y_1^{p_1} \cdots \partial y_d^{p_d}}, \, p \in \mathcal{P}_\beta \right),$$

1^0. *Deviations of $\hat{M}$.* By definition of $\hat{M}$ in (3.3), we obtain

$$|\hat{M} - M(f)| \leq ||\tilde{\delta} - \mathcal{D}(f)||_1 \leq \|\mathcal{V} \, F(X) - \mathcal{D}(f)\|_1 + \|\mathcal{V} \, V(X)\|_1 \, .$$

Here $\mathcal{V}$ is $D_b \times n$-matrix of general term $\mathcal{V}_{pi} = W_{ni}^p(y)$ and $\|.\|_1$ is the ℓ_1-norm. Let us prove that

$$\mathbb{P}_f\left\{|\hat{M} - M(f)| > M(f)/2\right\} \leq \exp\left\{-\frac{n^{\frac{b}{b+d}}}{8\vartheta_2^2 D_b^2}\right\}. \tag{7.19}$$

In view of the result proved in [6] and [21] there exist $\vartheta_1, \vartheta_2 > 0$ such that

$$\begin{aligned} \|\mathcal{V}\, F(X) - \mathcal{D}(f)\|_1 &\leq \vartheta_1 h_{\max}^{\beta - \lfloor \beta \rfloor}, \\ \sup_{i,x} |W_{ni}^p(y)| &\leq \frac{\vartheta_2}{n h_{\max}^d}, \quad p \in \mathcal{P}_\beta. \end{aligned}$$

Remind that $h_{\max} \xrightarrow{n\to\infty} 0$ and, therefore, $\exists n_0$ such that $\vartheta_1 h_{\max}^{\beta - \lfloor \beta \rfloor} \leq M(f)/4$ for any $n \geq n_0$. Note that n_0 can be chosen independently of f since $M(f)/4 \geq A/4$. Thus, we get

$$\begin{aligned} &\mathbb{P}_f\left\{|\hat{M} - M(f)| > M(f)/2\right\} \\ &\leq \sum_{p\in\mathbb{N}^d:\, 0\leq|p|\leq\beta} \mathbb{P}_f\left\{\left|\sum_{X_i\in[0,1]^d} f(X_i)(2U_i - 1)W_{ni}^p(y)\right| > \frac{M(f)}{4D_b}\right\}. \end{aligned}$$

Noting that $\left|f(X_i)(2U_i - 1)W_{ni}^p(y)\right| \leq M(f)\dfrac{\vartheta_2}{n h_{\max}^d}$, applying Höeffding inequality and the last inequality, we obtain

$$\begin{aligned} &\sum_{p\in\mathbb{N}^d:\, 0\leq|p|\leq\beta} \mathbb{P}_f\left\{\left|\sum_{X_i\in[0,1]^d} f(X_i)(2U_i - 1)W_{ni}^p(y)\right| > \frac{M(f)}{4D_b}\right\} \\ &\qquad \leq D_b \exp\left\{-\frac{n h_{\max}^d}{8\vartheta_2^2 D_b^2}\right\} = D_b \exp\left\{-\frac{n^{\frac{b}{b+d}}}{8\vartheta_2^2 D_b^2}\right\}. \end{aligned} \tag{7.20}$$

Therefore (7.19) is proved.

2^0. *Deviations of* $\hat{A}$. Since $|f(y) - A(f)| \leq L d h_{\max}^\beta \leq A(f)/4$ for $n \geq n_0$ one has

$$\mathbb{P}_f\left\{|\hat{A} - A(f)| > A(f)/2\right\} \leq \mathbb{P}_f\left\{|\hat{M} - M(f)| > A(f)/4\right\}.$$

Repeating previous calculations we obtain

$$\begin{aligned} \mathbb{P}_f\left\{|\hat{A} - A(f)| > A(f)/2\right\} &\leq D_b \exp\left\{-\frac{\big[A(f)\big]^2 n^{\frac{b}{b+d}}}{16\big[M(f)\big]^2 \vartheta_2^2 D_b^2}\right\} \\ &\leq D_b \exp\left\{-\frac{A n^{\frac{b}{b+d}}}{16 M \vartheta_2^2 D_b^2}\right\}. \end{aligned} \tag{7.21}$$

Since $\mathbb{P}_f\big(G^c\big) \leq \mathbb{P}_f\big(G^c_{\hat{A}}\big) + \mathbb{P}_f\big(G^c_{\hat{M}}\big)$ the assertion of the lemma follows from (7.20) and (7.21). ∎

Acknowledgements

I would like to thank Oleg Lepski for his helpful remarks and comments to improve the presentation of the paper. I am extremely grateful to two referees and the Associate Editor for their very helpful comments which are greatly appreciated.

References

[1] G. Allon, M. Beenstock, S. Hackman, U. Passy, and A. Shapiro. Nonparametric estimation of concave production technologies by entropic methods. *J. Appl. Econometrics*, 22(4):795–816, 2007.
[2] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.
[3] A. Goldenshluger and O. Lepski. Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14(3):1150–1190, 2008.
[4] A. Goldenshluger and O. Lepski. Structural adaptation via lp-norm oracle inequalities. *Probab. Theory and Related Fields*, 143:41–71, 2009.
[5] R.Z. Has'minskii and I.A. Ibragimov. *Statistical Estimation, Asymptotic Theory*. Springer-Verlag, Applications of Mathematics, 1981.
[6] W. Härlde, J. Hart, J.S. Marron, and A. Tsybakov. Bandwidth choice for average derivative estimation. *journal of the American Statistical Association*, 87:417, 1992.
[7] A. Juditsky, O. Lepski, and A. Tsybakov. Nonparametric estimation of composite functions. *Ann. Statist.*, 37(3):1360–1404, 2009.
[8] V. Katkovnik and V. Spokoiny. Spatially adaptive estimation via fitted local likelihood techniques. *IEEE Trans. Image Process.*, 56(3):873 –886, 2008.
[9] N. Klutchnikoff. *On the adaptive estimation of anisotropic functions*. PhD thesis, Aix-Masrseille 1, 2005.
[10] A.P. Korostelëv and A. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993. ISBN 0-387-94028-6.
[11] O. Lepski. On a problem of adaptive estimation in gaussian white noise. *Theory of Probability and its Applications*, 35(3):454–466, 1990.
[12] O. Lepski. Asymptotically minimax adaptive estimation i. upper bounds. optimally adaptive estimates. *Theory Probab. Appl.*, 36:682–697, 1991.
[13] O. Lepski. Asymptotically minimax adaptive estimation ii. statistical models without optimal adaptation. adaptive estimators. *Theory of Probability and its Applications*, 37:433–468, 1992.
[14] O. Lepski and V. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Annals of statistics*, 25(6):2512–2546, 1997.
[15] O. Lepski, E. Mammen, and V. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3):929–947, 1997. ISSN 0090-5364.

[16] A. Nemirovski. *Topics in non-parametric statistics*, volume 1738 of *Lecture Notes in Math.* Springer, Berlin, 2000.

[17] J. Polzehl and V. Spokoiny. Propagation-separation approach for local likelihood estimation. *Probab. Theory Related Fields*, 135(3):335–362, 2006.

[18] L. Simar and P. Wilson. Statistical inference in nonparametric frontier models: The state of the art. *Journal of Productivity Analysis*, 13:49–78, 2000.

[19] V. Spokoiny. Adaptive hypothesis testing using wavelets. *Ann. Statist.*, 24 (6):2477–2498, 1996.

[20] A. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of function on the sobolev classes. *Annals of statistics*, 26(6):2420–2469, 1998.

[21] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer Publishing Company, Incorporated, 2008.