SOME UNIFORM LIMIT RESULTS IN ADDITIVE REGRESSION MODEL

Mohammed DEBBARH

Université Paris 6
175, Rue du Chevaleret, 75013 Paris.
debbarh@ccr.jussieu.fr.

Key Words: additive model; curse of dimensionality; non parametric multivariate regression; marginal integration.

ABSTRACT

We establish some uniform limit results in the setting of additive regression model estimation. Our results allow to give an asymptotic 100% confidence bands for these components. These results are stated in the framework of i.i.d random vectors when the marginal integration estimation method is used.

arXiv:0706.1161v1 [math.ST] 8 Jun 2007

1 Introduction

For $d \geq 2$, let $(\mathbf{X}, Y)$ be an $\mathbb{R}^d \times \mathbb{R}$-valued random vector. The regression function of Y given $\mathbf{X} = \mathbf{x}$ is defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x}). \tag{1}$$

For $0 < p < \infty$, the optimal L_p rate of convergence of a nonparametric estimate of m is of order $n^{\frac{-k}{2k+d}}$ when m is assumed to be a k-times differentiable function and for $p = \infty$, the optimal rate is $(n^{-1}\log n)^{\frac{k}{2k+d}}$ (see, Stone (1985)). This rate of convergence which depends on the dimension d of the covariable $\mathbf{X}$ becomes worse as the dimensionality of the problem increases. In the literature, this phenomena is known under the name of "curse of dimensionality". To reduce the dimension impact upon the estimates, Stone (1985) proposed several sub-models of model (1). More particularly, he studied the nonparametric additive regression model in which the multivariate regression function is written as the sum of univariate functions, i.e,

$$m(\mathbf{x}) := m_{add}(\mathbf{x}) = \mu + \sum_{l=1}^{d} m_l(x_l). \tag{2}$$

To study the model (2), several estimation methods have been proposed, in the literature. We cite, the method based on B-spline (see, Stone (1985)), the method based on the backfitting algorithm (see, Hastie and Tibshirani (1986)); hereafter, we make use of the marginal integration method, (see, Newey (1994), Tjøstheim and Auestad (1994)and Linton and Nielsen (1995)). The additive regression components have motivated the work of many researchers, we refer to Camlong-Viot *et al.* (2000) for a survey on the asymptotic normality of the additive components under a mixing condition and to Sperlich *et al.* (2002) for nonparametric estimation and testing of integration in additive models. Debbarh (2004) established the law of iterated logarithm for additive regression model components under the independence assumption on the sample $(\mathbf{X}_i, Y_i)_{i=1,...,n}$. In this paper, we establish some uniform limit results in probability in an additive regression model. Similarly to results stated by Deheuvels and Mason (2004) for functionals of a distribution based on kernel-type estimation, our results allow us to build asymptotic 100% confidence bands for the components we estimate.

2 Main results

First, introduce some notations and assumptions. We denote by f and f_l the joint density of $\mathbf{X}$ and the marginal density of X_l, for $l = 1, ..., d$, respectively. We consider the following assumptions upon m, f and $f_l, l = 1, ..., d$. The functions f and f_l are continuous with compact supports and there exist b, b_l, B, and B_l such that

$(F.1)$ $0 < b \leq f(\underline{\mathrm{x}}) \leq B < \infty$ and $0 < b_l \leq f_l(x_l) \leq B_l < \infty$.
$(F.2)$ m is k-times continuously differentiable.
$(F.3)$ f is k'-times continuously differentiable, $k' > kd$.

Throughout, $(\ell_n)_{n\geq1}$, $(h_n)_{n\geq1}$ and $(h_{l,n})_{n\geq1}$, $1 \leq l \leq d$, are sequences of positive constants satisfying the following conditions

$(H.1)$ $h_{l,n} \to 0$, and $\ell_n \to 0$ as $n \to \infty$.
$(H.2)$ $nh_{1,n} \to +\infty$, $nh_{1,n}/\log(n) \to \infty$ as $n \to \infty$.
$(H.3)$ $nh_{1,n}h_{1,n}^{2k_1}...h_{d,n}^{2k_d}/|\log h_{1,n}| \longrightarrow 0$, $k_1 + ... + k_d = k$, and $h_{1,n}\log(n)/\ell_n^d|\log h_{1,n}| \longrightarrow 0$ as $n \to \infty$.
$(H.4)$ $h_n \sim n^{-2k/(2k+1)}$ and $nh_{1,n}h_n^2/|\log h_{1,n}| \to 0$ as $n \to \infty$.

Let $I = \prod_{i=1}^d I_i = \prod_{i=1}^d [a_i, c_i]$ and $J = \prod_{i=1}^d J_i = \prod_{i=1}^d [a'_i, c'_i]$ be two fixed pavements of

$\mathbb{R}^d$ such that $a_i' < a_i < c_i < c_i'$, $1 \le i \le d$. Furthermore, we consider the following assumptions upon the random variable Y.

$(M.1)$ $Y\mathbb{I}_{\{\mathbf{X}\in J\}}$ is bounded.

Let $(\mathbf{X}_i, Y_i)_{i=1,\dots,n}$ be a n-sample with the same distribution as $(\mathbf{X}, Y)$. Let L be a kernel on $\mathbb{R}^d$, of order k', bounded and with compact support. We define the kernel estimator $\widehat{f}_n$ of the density f by

$$\widehat{f}_n(\mathbf{x}) = \frac{1}{n\ell_n^d}\sum_{i=1}^{d} L\Big(\frac{\mathbf{x}-\mathbf{X}_i}{\ell_n}\Big).$$

To estimate the multivariate regression function defined in (1), we will be used the tow following estimators,

$$\widehat{m}_n(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i}{n\widehat{f}_n(X_i)}\Big(\prod_{l=1}^{d}\frac{1}{h_{l,n}}K_l\Big(\frac{x_l - X_{i,l}}{h_{l,n}}\Big)\Big), \tag{3}$$

and

$$\widetilde{m}_n(\mathbf{x}) = \sum_{i=1}^{n} \frac{Y_i}{nh_n^d\widehat{f}_n(X_i)}K\Big(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\Big), \tag{4}$$

where, the kernel functions K and K_l, $l = 1, ..., d$, are bounded, continuous, and integrate to one. In addition, we assume that K_l satisfies the following conditions

$(K.1)$ $\mathbb{K} := \prod_{l=1}^{d} K_l$ is of order k.

$(K.2)$ $K_l(u) = 0$ for $u \notin [-\frac{\lambda_l}{2}, \frac{\lambda_l}{2})$, for some $0 < \lambda_l < \infty$.

$(K.3)$ $\mathbb{K} = \prod_{l=1}^{d} K_l$ is square integrable function in the linear span (the set of finite linear combinations) of functions $\Psi \ge 0$ satisfying the following property: the subgraph of Ψ, $\{(s,u) : \Psi(s) \ge u\}$, can represented as a finite number of Boolean operations among sets of the form $\{(s,u) : p(s,u) \ge \phi(u)\}$, where p is a polynomial on $\mathbb{R}^d \times \mathbb{R}$ and ϕ arbitrary real function.
In particular this is satisfied by $K(\mathbf{x}) = \phi(p(\mathbf{x}))$, p being a polynomial and ϕ a bounded real function of bounded variation (see Nolan and Pollard (1987) and Giné and Guillou (2002)).

As already mentioned, the marginal integration method will be used to estimate the additive components (see Linton and Nielsen (1995) and Newey (1994)). Towards this aim, for all $\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d$ and every $\mathbf{x}_{-l} = (x_1, .., x_{l-1}, x_{l+1}, .., x_d)$, $l = 1, ..., d$, set $q(\mathbf{x}) = \prod_{l=1}^{d} q_l(x_l)$ and $q_{-l}(\mathbf{x}_{-l}) = \prod_{j \neq l} q_j(x_j)$. Then, the l-th component η_l of the additive model is given by

$$\eta_l(x_l) = \int_{\mathbb{R}^{d-1}} m(\mathbf{x}) q_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} m(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \quad l = 1, ..., d, \tag{5}$$

in such a way that the two following equalities hold,

$$\eta_l(x_l) = m_l(x_l) - \int_{\mathbb{R}} m_l(z) q_l(z) dz, \quad l = 1, ..., d, \tag{6}$$

$$m(\mathbf{x}) = \sum_{l=1}^{d} \eta_l(x_l) + \int_{\mathbb{R}^d} m(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}. \tag{7}$$

In view of (6) and (7), η_l and m_l are equal up to an additive constant. Therefore, η_l is an additive component too, fulfilling a different identifiability condition. Note also that $\eta_l = m_l$ for the particular choice $q_l = f_l$, $l = 1, ..., d$. However, f_l is generally unknown, and $\eta_l \neq m_l$ in practice. From (3) and (5), a natural estimate of the l-th component η_l is given by

$$\widehat{\eta}_l(x_l) = \int_{\mathbb{R}^{d-1}} \widehat{m}_n(\mathbf{x}) q_{-l}(\mathbf{x}_{-l}) d\mathbf{x}_{-l} - \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}, \; l = 1, ..., d. \tag{8}$$

From (7) and (8), we derive an estimate $\widehat{m}_{add}$ of the additive regression function,

$$\widehat{m}_{add}(\mathbf{x}) \;\; = \;\; \sum_{l=1}^{d} \widehat{\eta}_l(x_l) + \int_{\mathbb{R}^d} \widetilde{m}_n(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}. \tag{9}$$

In other respects, we impose the following assumptions on the known integration density functions q_{-l} and q_l, $l = 1, ..., d$,

$(Q.1)$ q_{-l} is bounded and continuous, $l = 1, ..., d$.
$(Q.2)$ q_l has k continuous and bounded derivatives, with compact support $\mathcal{C}_l \subset I_l$, $l = 1, ..., d$.

Let $\phi(u_1)$ be a continuous function on the interval I_1 defined by

$$\phi(u_1) = \int_{\mathbb{R}^{d-1}} \frac{H(\mathbf{u})}{f(\mathbf{u}_{-1}|u_1)} q_{-1}(\mathbf{u}_{-1}) d\mathbf{u}_{-1},$$

where

$$H(\mathbf{u}) = E(Y^2|\mathbf{X} = \mathbf{u}), \quad \mathbf{u} = (u_1, ..., u_d) \in \mathbb{R}^d.$$

Consider the following quantity

$$\sigma_l = \sup_{x_l \in I_l} \sqrt{\frac{\phi(x_l)}{f_l(x_l)} \int_{\mathbb{R}} K_l^2}. \tag{10}$$

The following results describe the asymptotic behavior of the estimates $\widehat{\eta}_1$ and $\widehat{m}_{add}$. From now on, $\overset{P}{\rightarrow}$ denote the convergence in probability.

Theorem 1 *Under the hypotheses* $(F.1)-(F.3)$, $(H.1)-(H.4)$, $(K.1)-(K.3)$, $(Q.1)-(Q.2)$ *and* $(M.1)$, *we have, as* $n \longrightarrow \infty$,

$$\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}} \sup_{x_1 \in I_1} \pm\{\widehat{\eta}_1(x_1) - \eta_1(x_1)\} \overset{P}{\longrightarrow} \sigma_1. \tag{11}$$

Theorem 2 below is valid under the additional condition that for any all $1 \leq l \leq d$, $h_{l,n} = h_{1,n}$.

Theorem 2 *Under the hypotheses* $(F.1)-(F.3)$, $(H.1)-(H.4)$, $(K.1)-(K.3)$, $(Q.1)-(Q.2)$ *and* $(M.1)$, *we have, as* $n \longrightarrow \infty$,

$$\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}} \sup_{\mathbf{x} \in I} \pm\Big\{\widehat{m}_{add}(\mathbf{x}) - m(\mathbf{x})\Big\} \overset{P}{\longrightarrow} \sum_{l}^{d} \sigma_l.$$

3 Application

3.1 Confidence bands

Let $\sigma_{1,n}(x_1)$ be the estimator of $\sigma_1(x_1)$, with $\sigma_1(x_1) = \sqrt{\phi(x_1)/f_1(x_1)}$. We will consider the data-dependent function $L_n(x_1)$, defined by,

$$L_n(x_1) = \Big\{\frac{2\log(1/h_{1,n})}{nh_{1,n}} \times \sigma_{1,n}(x_1)\Big\}^{1/2} \Big[\int_{\mathbb{R}} K_1^2\Big]^{1/2},$$

where

$$\sigma_{1,n}^2(x_1) = \frac{1}{nh_{1,n}} \sum_{i=1}^{n} Y_i^2 K_1\Big(\frac{x_1 - X_{i,1}}{h_{1,n}}\Big) \int_{\mathbb{R}^{d-1}} \frac{\prod_{l \neq 1} \frac{1}{h_{l,n}} K_l\Big(\frac{x_l - X_{i,l}}{h_{l,n}}\Big)}{\widehat{f}_n^2(\mathbf{x})} q_{-1}(\mathbf{x}_{-1})\, d\mathbf{x}_{-1}.$$

We obtain asymptotic simultaneous confidence bands for $\eta_1(x_1)$ in the following sense. For each $0 < \epsilon < 1$, we have

$$\mathbb{P}\Big\{\eta_1(x_1) \in \Big[\widehat{\eta}_1(x_1) - (1+\epsilon)L_n(x_1),\ \ \widehat{\eta}_1(x_1) + (1+\epsilon)L_n(x_1)\Big],\ \forall x_1 \in I_1\Big\} \longrightarrow 1,$$

and

$$\mathbb{P}\Big\{\eta_1(x_1) \in \Big[\widehat{\eta}_1(x_1) - (1-\epsilon)L_n(x_1), \quad \widehat{\eta}_1(x_1) + (1-\epsilon)L_n(x_1)\Big], \ \forall x_1 \in I_1\Big\} \longrightarrow 0.$$

We say then that the intervals

$$\Big[A_{n,1}(x_1), \quad B_{n,1}(x_1)\Big] = \Big[\widehat{\eta}_1(x_1) - L_n(x_1), \quad \widehat{\eta}_1(x_1) + L_n(x_1)\Big],$$

provide asymptotic simultaneous confidence bands (at an asymptotic confidence level of 100 %) for $\eta_1(x_1)$ over $x_1 \in I_1$. We deduce the asymptotic confidence bands for m_{add}, over $\mathbf{x} \in I$,

$$\Big[A_n(\mathbf{x}), \quad B_n(\mathbf{x})\Big] = \Big[\mu_n + \sum_{l=1}^{d} A_{n,l}(x_l), \quad \mu_n + \sum_{l=1}^{d} B_{n,l}(x_l)\Big], \text{ where } \mu_n = \int_{\mathbb{R}^d} \widehat{m}_n(\mathbf{x})q(\mathbf{x})d\mathbf{x}.$$

Following a general statistical practice, for finite values of the sample size n, we recommend the use of the asymptotic 100 % confidence bands. Our results do not provide confidence regions in the usual sense, since they are not based on a specified level $1-\alpha$. Instead, they hold with probability tending to 1 as $n \to \infty$.

4 Proofs

Let $\mathcal{G}$ be a class of pointwise measurable functions satisfying conditions $(\mathcal{C})$ in the Appendix. We denote by $\alpha_n(.)$ the multivariate empirical processus based upon $(\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), ...$ and indexed by the class of functions $\mathcal{G}$. More precisely, α_n is defined for $g \in \mathcal{G}$ by,

$$\alpha_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(g(\mathbf{X}_i, Y_i) - E(g(\mathbf{X}_i, Y_i))\Big).$$

For any real valued function ϕ defined on a set B, we use the notation $||\phi||_B = \sup_{x\in B} |\phi(x)| := ||\phi||$. Recalling that $I_1 = [a_1, c_1]$, let $0 < \eta < 1$ be a fixed number and set, for $n \geq 1$,

$$x_{1,j} = a_1 + j\eta h_{1,n}, \quad 0 \leq j \leq l_n := \Big[\frac{c_1 - a_1}{\eta h_{1,n}}\Big], \tag{12}$$

where $[u]$ denotes the integer part of u. For $\mathbf{X}_i = (X_{i,1}, ..., X_{i,d})$, $1 \leq i \leq n$, set

$$g_n^{x_1}(\mathbf{X}_i, Y_i) = Y_i G(\mathbf{X}_i) K_1\Big(\frac{x_1 - X_{i,1}}{h_{1,n}}\Big), \tag{13}$$

where,

$$G(\mathbf{X}_i) = \frac{1}{f(\mathbf{X}_i)} \int_{\mathbb{R}^{d-1}} \Big(\prod_{l\neq 1} \frac{1}{h_{1,n}} K_l\Big(\frac{x_l - X_{i,l}}{h_{l,n}}\Big) q_l(x_l)\Big) d\mathbf{x}_{-l}. \tag{14}$$

For $n \geq 1$ and any $0 \leq j \leq l_n$, let $\mathcal{G}_n = \{g_n^{x_{1,j}} : 0 \leq j \leq l_n\}$. Obviously, for each $0 \leq j \leq l_n$ and any $x_1 \in I_1$, we have

$$||g_n^{x_{1,j}}|| + ||g_n^{x_1}|| \leq \kappa, \quad \text{where} \quad \kappa \text{ is a positive constant.}$$

In the first time, we assume that the density f of $\mathbf{X}$ is known. Let $\widehat{\widehat{m}}_n$ be the estimator of the regression function when f is known,

$$\widehat{\widehat{m}}_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{Y_i}{f(\mathbf{X}_i)}\Big(\prod_{l=1}^{d} \frac{1}{h_{l,n}} K_l\Big(\frac{x_l - X_{i,l}}{h_{l,n}}\Big)\Big).$$

Using the estimator (8) of the additive regression model components, we obtain

$$\begin{aligned} nh_{1,n}\Big(\widehat{\widehat{\eta}}_1(x_1) - E(\widehat{\widehat{\eta}}_1(x_1))\Big) &= nh_{1,n}\int_{\mathbb{R}^{d-1}} \Big(\widehat{\widehat{m}}_n(\mathbf{x}) - E(\widehat{\widehat{m}}_n(\mathbf{x}))\Big) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1} \\ &\quad -nh_{1,n}\int_{\mathbb{R}^{d}} \Big(\widehat{\widehat{m}}_n(\mathbf{x}) - E(\widehat{\widehat{m}}_n(\mathbf{x}))\Big) q(x) d\mathbf{x}, \\ &= \sqrt{n}\alpha_n(g_n^{x_1}) - \int_{\mathbb{R}} \sqrt{n}\alpha_n(g_n^{x_1}) q_1(x_1) dx_1, \end{aligned}$$

where,

$$\sqrt{n}\alpha_n(g_n^{x_1}) = nh_{1,n}\int_{\mathbb{R}^{d-1}} \Big(\widehat{\widehat{m}}_n(\mathbf{x}) - E(\widehat{\widehat{m}}_n(\mathbf{x}))\Big) q_{-1}(\mathbf{x}_{-1}) d\mathbf{x}_{-1}. \tag{15}$$

The proof of Theorem 1 is based on a number of additional lemmas.

Lemma 1 *Assume that the conditions* $(F.1)-(F.3)$, $(H.1)-(H.3)$, $(K.1)-(K.3)$, $(Q.1)-(Q.2)$ *and* $(M.1)$, *are satisfied. Then, we have, as* $n \longrightarrow \infty$

$$\frac{\sup_{x_1 \in I_1} \pm\alpha_n(g_n^{x_1})}{\sqrt{2h_{1,n}|\log h_{1,n}|}} \xrightarrow{P} \sigma_1.$$

Following Deheuvels and Mason (2004) and Einmahl and Mason (2000), the proof of Lemma 1 is split up into two part. First, we establish the upper bound, afterward, we state the lower bound.

4.1 Proof of Lemma 1

4.1.1 Upper bound part

The main tools used in the proof are the discretization and the properties of empirical process oscillations.

Part 1. We examine the behavior of our process on an appropriately chosen grid, $(x_{1,j})_{1\leq j\leq l_n}$ of I_1. Assume that the assumptions of Lemma 1 hold. It follows from (13) that

$$\begin{aligned} Var\Big(g_n^{x_{1,j}}(\mathbf{X}_i,Y_i))\Big) &\leq E\Big(\Big(g_n^{x_{1,j}}(\mathbf{X}_i,Y_i)\Big)^2\Big) \\ &\leq E\Big(G^2(\mathbf{X}_i)Y_i^2K_1^2\Big(\frac{x_{1,j}-X_{i,1}}{h_{1,n}}\Big)\Big) \\ &\leq E\Big(E(Y_i^2|\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\Big(\frac{x_{1,j}-X_{i,1}}{h_{1,n}}\Big)\Big) \\ &\leq E\Big(H(\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\Big(\frac{x_{1,j}-X_{i,1}}{h_{1,n}}\Big)\Big). \end{aligned} \quad (16)$$

But, making use of classical changes of variables and Taylor expansion, we get under $(K.1)$ with $\mathbf{h}_{-1}=(h_{2,n},...,h_{d,n})^T$ and $0<\theta<1$,

$$\begin{aligned} &\int_{\mathbb{R}^{d-1}}\prod_{l\neq 1}\Big(\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-u_l}{h_{l,n}}\Big)q_l(x_l)\Big)d\mathbf{x}_{-1} \\ = &\int_{\mathbb{R}^{d-1}}\prod_{l\neq 1}\Big(K_l(v_l)q_l(v_lh_{l,n}+u_l)\Big)dx_{-1} \\ = &\int_{\mathbb{R}^{d-1}}\Big(\prod_{l\neq 1}K_l(v_l)\Big)q_{-1}(\mathbf{v}_{-1}\mathbf{h}_{-1}+\mathbf{u}_{-1})d\mathbf{v}_{-1} \\ = &\int_{\mathbb{R}^{d-1}}\Big(\prod_{l\neq 1}K_l(v_l)\Big)\Big[q_{-1}(\mathbf{u}_{-1})+\sum_{k_2+...+k_d=k}v_2^{k_2}...v_d^{k_d}h_{2,k}^{k_2}...h_{d,k}^{k_d}\frac{\partial^k q_{-1}}{\partial v_2^{k_1}...\partial v_d^{k_d}}(\mathbf{v}_{-1}\mathbf{h}_{-1}\theta+\mathbf{u}_{-1})\Big]d\mathbf{v}_{-1} \\ = &\ q_{-1}(\mathbf{u}_{-1})+o(1). \end{aligned}$$

Therefore, it hold that,

$$\Big(\int_{\mathbb{R}^{d-1}}\prod_{l\neq 1}\Big(\frac{1}{h_{1,n}}K_l\Big(\frac{x_l-u_l}{h_{l,n}}\Big)q_l(x_l)\Big)d\mathbf{x}_{-1}\Big)^2=q_{-1}^2(\mathbf{u}_{-1})+o(1). \quad (17)$$

Moreover, in view of the (14), we have

$$\begin{aligned} &E\Big(H(\mathbf{X}_i)G^2(\mathbf{X}_i)K_1^2\Big(\frac{x_{1,j}-X_{i,1}}{h_{1,n}}\Big)\Big) \\ = &\int_{\mathbb{R}^d}\frac{H(\mathbf{u})}{f(\mathbf{u})}\Big(\int_{\mathbb{R}^{d-1}}\prod_{l\neq 1}\frac{1}{h_{1,n}}K_l\Big(\frac{x_l-u_l}{h_{l,n}}\Big)q_l(x_l)dx_{-1}\Big)^2K_1^2\Big(\frac{x_{1,j}-u_1}{h_{1,n}}\Big)d\mathbf{u}, \end{aligned} \quad (18)$$

Combining (16), (17) and (18), we obtain

$$Var\Big(g_n^{x_{1,j}}(\mathbf{X}_i,\mathbf{Z}_i,\delta_i)\Big)\leq \sigma_1^2h_{1,n}+o(h_{1,n}). \quad (19)$$

Applying Bernstein's inequality to the sequence of random variables,

$$Z_r = g_n^{x_{1,j}}(\mathbf{X}_r, Y_r) - E(g_n^{x_{1,j}}(\mathbf{X}_r, Y_r)), \quad r = 1, ..., n,$$

we obtain, for n large enough, that,

$$\begin{aligned} & \mathbb{P}\Big\{ \max_{1 \le j \le l} |\alpha_n(g_n^{x_{1,j}})| \ge \sigma_1(1+\rho)\sqrt{2h_{1,n}|\log h_{1,n}|} \Big\} \\ \le \;& 2(l+1)\exp\left(\frac{-2\sigma_1^2(1+\rho)^2 h_{1,n}|\log h_{1,n}|}{2\sigma_1^2 h_{1,n} + \frac{2M}{3\sqrt{n}}\sigma_1\sqrt{2h_{1,n}|\log h_{1,n}|}} \right) \\ \le \;& 2(l+1) h_{1,n}^{1+\frac{\rho}{2}}. \end{aligned} \tag{20}$$

Part 2. Under assumption $(K.3)$, the class of functions

$$\Big\{ K\Big(\frac{\mathbf{t} - .}{a}\Big) : t \in \mathbb{R}^d, a \in \mathbb{R}^d \setminus \{0\} \Big\}$$

is a bounded VC class of measurable functions. Now, consider the class

$$\mathcal{F} = \Big\{ byK\Big(\frac{\mathbf{t} - .}{a}\Big) : t \in \mathbb{R}^d, a \in \mathbb{R}^d \setminus \{0\}, |b| \le D \Big\},$$

where $D > 0$ is the bound of the function $yG(\mathbf{x})$. Arguing exactly as in pages 254 and 255 of Deheuvels and Mason (2004), one can show that $\mathcal{F}$ fulfills $\mathcal{C}$. An easy argument now shows that

$$\mathcal{G}' = \Big\{ byK\Big(\frac{\mathbf{t} - .}{a}\Big) - b'yK\Big(\frac{\mathbf{t}' - .}{a'}\Big) : \mathbf{t}, \mathbf{t}' \in \mathbb{R}^d, a, a' \in \mathbb{R} \setminus \{0\}, |b| \le D, |b'| \le D \Big\}$$

fulfills conditions $\mathcal{C}$. As over $\mathbb{G}$ function we can take

$$\mathbb{G}(\mathbf{u}, v) = 2Cv||K||_\infty.$$

Since $\mathcal{G}'_n \subset \mathcal{G}'$. We study the behavior of our process between the grid points $x_{1,j}, \; x_{1,j+1}$, with $1 \le j \le l_n$. Toward this aim, consider for $0 \le j \le l_n$, the class of functions

$$\mathcal{G}'_{n,j} = \{ g_n^{x_{1,j}} - g_n^{x_1}, \; x_{1,j} \le x_1 \le x_{1,j+1} \} \;\text{ and }\; \mathcal{G}'_n = \bigcup_j \mathcal{G}'_{n,j}.$$

There exists an absolute constant B, such that for any $\epsilon > 0$, one can find a η_ϵ such that whenever (12) holds, with $0 < \eta < \eta_\epsilon$, we have,

$$\mathbb{P}\Big\{ ||n^{1/2}\alpha_n||_{\mathcal{G}'_n} \ge B\sqrt{\epsilon n h_{1,n}|\log h_{1,n}|} \Big\} = o(1). \tag{21}$$

Indeed, we see that, uniformly over $g \in \mathcal{G}'_{n,j} \subset \mathcal{G}'_n$, $||g|| \leq \kappa$. Moreover, by similar arguments as those used in the proof of (16), we have,

$$\sigma^2_{\mathcal{G}'_n} = \sup_{g\in\mathcal{G}'_n} \mathrm{Var}(g(\mathbf{X},Y)) \leq 4h_{1,n}\sigma_1^2. \tag{22}$$

Therefore by Fact 1 (see the Appendix), for all $t > 0$ we have for suitable finite constants $A_1, A_2 > 0$,

$$\begin{aligned} &\mathbb{P}\Big\{||n^{1/2}\alpha_n||_{\mathcal{G}'_n} \geq A_1\Big(E||\sum_{i=1}^n \epsilon_i g(\mathbf{X}_i,Y_i)||_{\mathcal{G}'_n} + t\Big)\Big\} \\ \leq\ & 2\Big\{\exp\Big(-\frac{A_2t^2}{n\sigma^2_{\mathcal{G}'_n}}\Big) + \exp\Big(-\frac{A_2t}{\kappa}\Big)\Big\} \\ \leq\ & 2\Big\{\exp\Big(-\frac{A_2t^2}{2nh_{1,n}\sigma_1^2}\Big) + \exp\Big(-\frac{A_2t}{\kappa}\Big)\Big\}. \end{aligned} \tag{23}$$

Next, by using (12), in combination with Fact 2 (see the Appendix), we obtain that

$$E||\sum_{i=1}^n \epsilon_i g_n^{x_1}(\mathbf{X}_i,Y_i)||_{\mathcal{G}'_n} \leq A'\sqrt{\nu n h_{1,n}\log(1/h_{1,n})}. \tag{24}$$

Where A' is an absolute constant. Thus, using (24), we get from (23) that

$$\mathbb{P}\Big\{||n^{1/2}\alpha_n||_{\mathcal{G}'_n} \geq 2A_1A'\sqrt{\nu n h_{1,n}|\log h_{1,n}|}\Big\} \tag{25}$$

$$\leq 2\exp\Big(-\frac{A_2(A_1A')^2\nu|\log h_{1,n}|}{\sigma_1^2}\Big) + 2\exp\Big(-\frac{A_2(A_1A')\sqrt{\nu n h_{1,n}|\log h_{1,n}|}}{M}\Big)$$

$$= o(1). \tag{26}$$

Taking $B = 2A_1A'\sqrt{\nu\epsilon}$ in (25), we complete the proof of (21).

For $1 \leq j \leq l_n$, $\mathcal{G}'_{n,j} \subseteq \mathcal{G}_n$, we see that

$$\frac{\max_{0\leq j\leq l}||n^{1/2}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_{1,n}|\log h_{1,n}|}} \leq \frac{||n^{1/2}\alpha_n||_{\mathcal{G}'_n}}{\sqrt{2nh_{1,n}|\log h_{1,n}|}}. \tag{27}$$

Using the statement (21) and the inequality (27) with $A = \frac{B}{\sigma_1\sqrt{2}}$, we obtain

$$\mathbb{P}\Big\{\frac{\max_{1\leq j\leq l_n}||n^{1/2}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_{1,n}|\log h_{1,n}|}} \geq \sigma_1 A\sqrt{\epsilon}\Big\} = o(1), \tag{28}$$

Conclusion : By combining (12) and (28), we conclude that there exists an absolute constant $A > 0$, such that

$$\mathbb{P}\Big\{\frac{\sup_{x_1\in I_1}|\alpha_n(g_n^{x_1})|}{\sqrt{2h_{1,n}|\log h_{1,n}|}} > (1+\rho+A\sqrt{\epsilon})\sigma_1\Big\} \leq \mathbb{P}\Big\{\max_{1\leq j\leq l}\frac{|\alpha_n(g_n^{x_{1,j}})|}{\sqrt{2h_{1,n}|\log h_{1,n}|}} > (1+\rho)\sigma_1\Big\}$$
$$+\mathbb{P}\Big\{\max_{1\leq j\leq l_n}\frac{||n^{\frac{1}{2}}\alpha_n||_{\mathcal{G}'_{n,j}}}{\sqrt{2nh_{1,n}|\log h_{1,n}|}} > A\sqrt{\epsilon}\sigma_1\Big\}.$$

Since for any $\epsilon > 0$, we can choose $\rho > 0$ and $\epsilon > 0$ small enough so that $\rho + A\sqrt{\epsilon} < \epsilon$. We obtain the upper bound result in the case where f is known,

$$\mathbb{P}\Big\{\frac{\sup_{x_1\in I_1}|\alpha_n(g_n^{x_1})|}{\sqrt{2h_{1,n}|\log h_{1,n}|}} > (1+\epsilon)\sigma_1\Big\} = o(1).$$

4.1.2 Lower bound part

In order to prove lower bound result, we gather hereafter some technical results (see, Einmahl and Mason (2000)). Let $Z_1, Z_2, ...$ be a sequence of i.i.d random vectors taking values in $\mathbb{R}^d$. For each $n \geq 1$, consider the empirical distribution function defined by,

$$G_n(\mathbf{s}) = n^{-1}\sum_{i=1}^{n}\mathbb{I}_{\{\mathbf{Z}_i\leq \mathbf{s}\}},\ \ s\in\mathbb{R}^{d+1},$$

where as usual $\mathbf{z} \leq \mathbf{s}$ means that each component of $\mathbf{z}$ is less than or equal to the corresponding component of s. For any measurable real valued function $g(.)$ defined on $\mathbb{R}^{d+1}$, set

$$G_n(g) = \int_{\mathbb{R}^d} g(\mathbf{s})\ dG_n(\mathbf{s}),\quad \mu(g) = E(g(\mathbf{Z}))\text{ and }\sigma(g) = \sqrt{Var(g(\mathbf{Z}))}.$$

Let $\{a_n : n \geq 1\}$ denote a sequence of positive constants converging to zero and satisfying the condition $|\log(a_n)|/\log\log(n) \to \infty$. For some sequence of integer number k_n, consider a sequence of sets of real valued measurable functions on $\mathbb{R}^{d+1}$, $\mathcal{G}_n = \{g_i^{(n)};\ i = 1, ..., k_n\}$, defined by the following conditions:

(a) $\mathbb{P}\Big(g_i^{(n)}(Z) \neq 0,\ g_i^{(n)}(Z) \neq 0\Big) = 0,\ \ \forall 1 \leq i \neq j \leq k_n$ and $\sum_{i=1}^{k_n} P\{g_i^{(n)}(Z) \neq 0) \leq 1/2$.

Furthermore, assume that for some $0 < r < \infty$,

(b) $a_nk_n \to r$ as $n \to \infty$.

For some $-\infty < \mu_1 < \mu_2 < \infty$ and $0 < \sigma_1 < \sigma_2 < \infty$, uniformly in $i = 1, ..., k_n$, we have for n large enough,

(c) $a_n\mu_1 \leq \mu(g_i^{(n)}) \leq a_n\mu_2$ and $\sqrt{a_n}\sigma_1 \leq \sigma(g_i^{(n)}) \leq \sqrt{a_n}\sigma_2$.

For some $0 < M_1 < \infty$, uniformly in $i = 1, ..., k_n$, we have for n large enough,

(d) $\quad ||g_i^{(n)}|| \leq M_1$.

The following lemma due to Einmahl and Mason (2000) is the main tool to prove our result. We will work only in the "+" case, the arguments for the "-" case can be obtained similarly.

Lemma 2 *Under assumptions* (**a**) – (**d**), *for each* $0 < \epsilon < 1$, *we have*

$$\mathbb{P}\Big\{ \max_{1 \leq i \leq k_n} \frac{n^{1/2}\{G_n(g_i^{(n)}) - \mu(g_i^{(n)})\}}{\sigma(g_i^{(n)})\sqrt{2|\log(a_n)|}} \geq 1 - \epsilon \Big\} \to 1.$$

Proof: See Proposition 2 of Einmahl and Mason (2000).

In order to apply the result of Lemma 2, we need to check the validity of the conditions (**a**) – (**d**) in our setting. For any $\epsilon > 0$, select a sub-interval $J_1 = [a_1', c_1']$ of $I_1 = [a_1, c_1]$, such that

$$\inf_{u_1 \in J_1} \sqrt{\frac{\phi(u_1)}{f_1(u_1)}} \Big[\int_{\mathbb{R}} K_1^2 \Big]^{1/2} > \sigma_1(1 - \epsilon/2)$$

and

$$\mathbf{P}\{X_1 \in J_1\} \leq 1/2.$$

Consider the following points in the interval J_1

$$x_{1,j} = a_1' + 2jh_{1,n}, \text{ for } j = 1, ..., [(b_1' - a_1')/2h_{1,n}] - 1 := k_n.$$

Then, it is easy to see that the condition (**b**) is satisfied with $a_n = h_{1,n}$, i.e.

$$\lim_{n \to \infty} h_{1,n} k_n \approx [(b_1 - a_1)/2].$$

For each $x_{1,j}$, $1 \leq j \leq k_n$, define the functions

$$g_j^{(n)}(\mathbf{x}, y) = yG(\mathbf{x})K_1\Big(\frac{x_{1,j} - x_1}{h_{1,n}}\Big),$$

it follows from $(F.1)$, that

$$||g_j^{(n)}|| \leq M||G|| \times ||K|| := M_1,$$

so the condition (**d**) is verified.

Now, we verify the validity of the condition (**a**). To this end, recall that K_1 is compactly supported. therefore,

$$g_i^{(n)}(\mathbf{X},Y) \;\neq\; 0 \Longleftrightarrow \Big|\frac{x_{1,i}-X_{i,1}}{h_{1,n}}\Big| \leq \frac{1}{2},$$

and then

$$|x_{1,j}-X_{i,1}| \;=\; |x_{1,j}-x_{1,i}+x_{1,i}-X_{i,1}| \geq 2h_{1,n}-\frac{h_{1,n}}{2}.$$

Hence , for $1 \leq i \neq j \leq k_n$,

$$\mathbb{P}\Big\{\; g_i^{(n)}(\mathbf{X},Y) \neq 0 \text{ and } g_j^{(n)}(\mathbf{X},Y) \neq 0\Big\} = 0.$$

To check the validity of the condition (**c**), recall the inequality (16) and observe that, for $x_1 \in J_1$,

$$\begin{aligned}
\mathrm{Var}(g_i^{(n)}(\mathbf{X},Y)) &= E\Big(Y_i^2G^2(X_i)K_1^2\Big(\frac{x_1-X_{i,1}}{h_{1,n}}\Big)\Big) - E\Big(Y_iG(X_i)K\Big(\frac{x_1-X_{i,1}}{h_{1,n}}\Big)\Big)^2 \\
&= \int_{\mathbb{R}} K_1^2\Big(\frac{x_1-u_1}{h_{1,n}}\Big)\frac{\phi(u_1)}{f_1(u_1)}du_1 + o(h_{1,n}) \\
&\geq h_{1,n}\Big(\int K_1^2\Big)\inf_{u_1\in J_1}\frac{\phi(u_1)}{f_1(u_1)} + o(h_{1,n}) \\
&\geq \sigma_1^2(1-\epsilon/2)h_{1,n}.
\end{aligned}$$

Now, we can use the result of Lemma 2, which yields,

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{1\leq i\leq k_n}\frac{n^{1/2}\{G_n(g_i^{(n)})-\nu(g_i^{(n)})\}}{\sqrt{2\mathrm{Var}(g_i^{(n)})|\log h_{1,n}|}} < 1-\epsilon\Big\} \\
\leq\;& \mathbb{P}\Big\{\max_{1\leq i\leq k_n}\frac{n^{1/2}(G_n(g_i^{n})-\nu(g_i^{(n)}))}{\sqrt{2h_{1,n}|\log h_{1,n}|}} < (1-\epsilon)^{3/2}\sigma_1 h_{1,n}^{1/2}\Big\}.
\end{aligned}$$

For $a_n = h_{1,n}$, we get from last inequality

$$\mathbb{P}\Big\{\max_{1\leq i\leq k_n}\frac{n^{1/2}(G_n(g_i^{n})-\nu(g_i^{(n)}))}{\sqrt{2h_{1,n}|\log h_{1,n}|}} < (1-\epsilon)^{3/2}\sigma_1 h_{1,n}^{1/2}\Big\} = o(1).$$

By setting $(1-\epsilon)^{3/2} = 1-\frac{\epsilon'}{2}$, and using the inequality,

$$\frac{\sup_{x_1\in I_1}\sqrt{n}\alpha_n(g_{n,1}^{x_1})}{\sqrt{2nh_{1,n}|\log h_{1,n}|}} \geq \max_{1\leq i\leq k_n}\frac{n^{1/2}(G_n(g_i^{n})-\nu(g_i^{(n)}))}{\sqrt{2h_{1,n}|\log h_{1,n}|}},$$

we obtain

$$\mathbb{P}\Big\{\frac{\sup_{x_1\in I_1}\alpha_n(g_n^{x_1})}{\sqrt{2h_{1,n},n|\log(h_{1,n})|}}<(1-\frac{\epsilon}{2})\sigma_1\Big\}=o(1).$$

Combining the results of Part 1 and Part 2, it follows that,

$$\Big|\frac{\sup_{x_1\in I_1}\alpha_n(g_n^{x_1})}{\sqrt{2h_{1,n}|\log h_{1,n}|}}-\sigma_1\Big|=o_{\mathbf{P}}(1).$$

Similarly one may show that

$$\Big|\frac{\sup_{x_1\in I_1}-\alpha_n(g_n^{x_1})}{\sqrt{2h_{1,n}|\log h_{1,n}|}}-\sigma_1\Big|=o_{\mathbf{P}}(1).$$

This completes the proof of Lemma 1.□

4.2 Proof of Theorem 1

Let us use the decomposition,

$$\begin{aligned}\pm\{\widehat{\eta}_1(x_1)-\eta_1(x_1)\} &= \pm\{\widehat{\eta}_1(x_1)-\widehat{\widehat{\eta}}_1(x_1)\}\pm\{\widehat{\widehat{\eta}}_1(x_1)-E(\widehat{\widehat{\eta}}_1(x_1))\}\pm\{E(\widehat{\widehat{\eta}}_1(x_1))-\eta_1(x_1)\}\\ &= T_1(x_1)+T_1(x_2)+T_3(x_3).\end{aligned}$$

First, consider the term $T_1(x_1)$. It hold that,

$$\widehat{m}_n(\mathbf{x})=\widehat{\widehat{m}}_n(\mathbf{x})+\frac{1}{n}\sum_{i=1}^{n}Y_i\Big(\prod_{i=1}^{d}\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-X_{i,l}}{h_{l,n}}\Big)\Big)\frac{f(\mathbf{X}_i)-\widehat{f}_n(\mathbf{X}_i)}{f(\mathbf{X}_i)\widehat{f}_n(\mathbf{X}_i)}.$$

It follows that,

$$\Big|\widehat{m}_n(\mathbf{x})-\widehat{\widehat{m}}_n(\mathbf{x})\Big| \leq \frac{1}{n}\sum_{i=1}^{n}|Y_i|\Big|\prod_{i=1}^{d}\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-X_{i,l}}{h_{l,n}}\Big)\Big|\times\frac{\sup_{1\leq i\leq n}\Big|f(\mathbf{X}_i)-\widehat{f}_n(\mathbf{X}_i)\Big|}{\Big|f(\mathbf{X}_i)\widehat{f}_n(\mathbf{X}_i)\Big|}.\tag{29}$$

Using, for example, the following result due to Ango-Nze and Rios (2000),

$$\sup_{\mathbf{x}\in I}|\widehat{f}_n(\mathbf{x})-f(\mathbf{x})|=O\Big(\sqrt{\frac{\log(n)}{n\ell_n^d}}\Big)\quad\text{a.s.},\tag{30}$$

we obtain under the assumptions $(F.1)$, $(M.1)$ and $(K.1)-(K.2)$, for n large enough,

$$\sup_{x\in I}|\widehat{m}_n(\mathbf{x})-\widehat{\widehat{m}}_n(\mathbf{x})|=O\Big(\sqrt{\frac{\log(n)}{n\ell_n^d}}\Big)\quad\text{a.s.}.\tag{31}$$

Observe that,

$$
\begin{aligned}
&\Big|\widehat{\widehat{\eta}}_1(x_1)-\widehat{\eta}_1(x_1)\Big| \\
&\quad = \Big|\int_{\mathbb{R}^{d-1}}\Big[\widehat{\widehat{m}}_n(\mathbf{x})-\widehat{m}_n(\mathbf{x})\Big]q_{-1}(\mathbf{x})d\mathbf{x}_{-1}+\int_{\mathbb{R}^d}\Big[\widehat{\widehat{m}}_n(\mathbf{x})-\widehat{m}_n(\mathbf{x})\Big]q(\mathbf{x})d\mathbf{x}\Big|, \\
&\quad \leq \int_{\mathbb{R}^{d-1}}\sup_{\mathbf{x}\in I}\Big|\widehat{\widehat{m}}_n(\mathbf{x})-\widehat{m}_n(\mathbf{x})\Big|q_{-1}(\mathbf{x}_{-1})dx_{-1}+\int_{\mathbb{R}^d}\sup_{\mathbf{x}\in I}\Big|\widehat{\widehat{m}}_n(\mathbf{x})-\widehat{m}_n(\mathbf{x})\Big|q(\mathbf{x})d\mathbf{x}, \\
&\quad \leq C\sqrt{\frac{\log(n)}{n\ell_n^d}} \qquad \text{a.s.},
\end{aligned}
$$

where C is a positive constant. Therefore, we have

$$
\sup_{x_1\in I_1}\mid\widehat{\eta}_1(x_1)-\widehat{\widehat{\eta}}_1(x_1)\mid= O\Big(\sqrt{\frac{\log(n)}{n\ell_n^d}}\Big) \quad \text{a.s..} \tag{32}
$$

Combining (32) and the assumption $(H.3)$, we obtain,

$$
\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}} \quad \sup_{x_1\in I_1}T_1(x_1)=o(1) \quad \text{a.s.} \tag{33}
$$

Next, turning our attention to $T_3(x_1)$. It hold that

$$
\begin{aligned}
\Big|m(\mathbf{x})-E(\widehat{m}_n(\mathbf{x}))\Big| &= \Big|m(\mathbf{x})-E\Big(\frac{Y_i}{f(\mathbf{X}_i)}\Big[\prod_{l=1}^d\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-X_{i,l}}{h_{l,n}}\Big)\Big]\Big)\Big| \\
&= \Big|m(\mathbf{x})-\int_{\mathbb{R}^d}m(\mathbf{u})\Big[\prod_{l=1}^d\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-u_l}{h_{l,n}}\Big)\Big]d\mathbf{u}\Big|.
\end{aligned}
$$

The change of variables and the Taylor expansion to order k, gives, with $0<\theta<1$ and $\mathbf{h}=(h_{1,n},...,h_{d,n})^T$

$$
\begin{aligned}
&\Big|m(\mathbf{x})-E(\widehat{m}_n(\mathbf{x}))\Big| \\
&\quad = \Big|m(\mathbf{x})-\int_{\mathbb{R}^d}m(-\mathbf{v}\mathbf{h}\theta+\mathbf{x})\Big(\prod_{i=1}^d K_i(v_i)\Big)d\mathbf{v}\Big| \\
&\quad \leq \int_{\mathbb{R}^d}|(m(\mathbf{x})-m(-\mathbf{v}\mathbf{h}\theta+\mathbf{x}))|\Big(\prod_{i=1}^d K_i(v_i)\Big)d\mathbf{v} \\
&\quad \leq \frac{1}{k!}\int_{\mathbb{R}^d}\sum_{k_1+...+k_d=k}h_{1,n}^{k_1}v_1^{k_1}...h_{d,n}^{k_d}v_d^{k_d}\Big|\frac{\partial^k m}{\partial v_1^{k_1}...\partial v_d^{k_d}}(-\mathbf{v}\mathbf{h}\theta+\mathbf{x})\Big|\Big(\prod_{i=1}^d K_i(v_i)\Big)d\mathbf{v} \\
&\quad \leq \frac{1}{k!}\Big\|\frac{\partial^k m}{\partial v_1^{k_1}...\partial v_d^{k_d}}\Big\|_\infty\sum_{k1+...+k_d=k}h_{1,n}^{k_1}...h_{d,n}^{k_d}\Big(\int_{\mathbb{R}}v_1^{k_1}...v_d^{k_d}\mathbb{K}(\mathbf{v})d\mathbf{v}\Big).
\end{aligned}
$$

It follows that,

$$\begin{aligned}\sup_{x_1\in I_1} T_3(x_1) &\le 2\sup_{\mathbf{x}\in I}\Big|m(\mathbf{x})-E(\widehat{m}_n(\mathbf{x}))\Big| \\ &\le \frac{2}{k!}\Big\|\frac{\partial^k m}{\partial v_1^{k_1}...\partial v_d^{k_d}}\Big\|_\infty \sum_{k_1+...+k_d=k} h_{1,n}^{k_1}...h_{d,n}^{k_d}\Big(\int_{\mathbb{R}} v_1^{k_1}...v_d^{k_d}\mathbb{K}(\mathbf{v})d\mathbf{v}\Big). \qquad (34)\end{aligned}$$

By combining the assumption $(H.3)$ and the statement (34), we obtain,

$$\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}}\sup_{x_1\in I_1} T_3(x_1)=o(1). \qquad (35)$$

Finally we evaluate the term $T_2(x_1)$. From (15), we observe that,

$$nh_1\{\widehat{\widehat{\eta}}_1(x_1)-E(\widehat{\widehat{\eta}}_1(x_1))\}=\sqrt{n}\alpha_n(g_n^{x_1})-\int_{\mathbb{R}}\sqrt{n}\alpha_n(g_n^{x_1})q_1(x_1)dx_1, \qquad (36)$$

then

$$\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}}\,T_2(x_1)=\frac{\pm\,\alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}}\mp\int_{\mathbb{R}}\frac{\alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}}q_1(x_1)dx_1.$$

We will use in this proof Camlong-Viot *et al.* (2000) notations,

$$\begin{aligned}\hat{\alpha}_1(x_1) &= \frac{1}{nh_{1,n}}\sum_{i=1}^{n}\frac{\tilde{Y}_{i,n}}{f_1(X_1)}K_1\Big(\frac{x_1-X_{i,1}}{h_{1,n}}\Big), \\ \tilde{Y}_{i,n} &= \int_{\mathbb{R}^{d-1}}\Big(\prod_{l=2}^{d}\frac{1}{h_{l,n}}K_l\Big(\frac{x_l-X_{i,l}}{h_{l,n}}\Big)\Big)\frac{q_{-1}(\mathbf{x}_{-1})}{f(X_{i,-1}|X_{i,1})}d\mathbf{x}.\end{aligned}$$

Using the Cauchy-Schwartz inequality, we obtain,

$$\begin{aligned}&\Big|\int_{\mathbb{R}}\frac{\alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}}\;q_1(x_1)dx_1\Big| \\ &= \sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}}\Big|\int_{\mathbb{R}^d}\{\widehat{m}_n(\mathbf{x})-E(\widehat{m}_n(\mathbf{x}))\}q(\mathbf{x})d\mathbf{x}\Big| \\ &= \sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}}\Big|\int_{\mathbb{R}}\{\hat{\alpha}_1(x_1)-E(\hat{\alpha}_1(x_1))\}q_1(x_1)dx_1\Big| \\ &\le \sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}}\Big[\int_{\mathcal{C}_1}\{\hat{\alpha}_1(x_1)-E(\hat{\alpha}_1(x_1))\}^2dx_1\Big]^{1/2}\times\Big[\int_{\mathbb{R}}q_1^2(x_1)dx_1\Big]^{1/2}. \qquad (37)\end{aligned}$$

By Camlong-Viot *et al.* (2000) we have, for all $x_1 \in \mathcal{C}_1$,

$$\mathrm{Var}(\hat{\alpha}_1(x_1)) = E\Big(\hat{\alpha}_1(x_1) - E(\hat{\alpha}_1(x_1))\Big)^2 = \mathcal{O}\Big(n^{-2k/(2k+1)}\Big),$$

then

$$\begin{aligned}\int_{\mathcal{C}_1} \mathrm{Var}(\hat{\alpha}_1(x_1))dx_1 &= \int_{\mathcal{C}_1} E\Big(\hat{\alpha}_1(x_1) - E(\hat{\alpha}_1(x_1))\Big)^2 dx_1 \\ &= E\Big(\int_{\mathcal{C}_1}\Big(\hat{\alpha}_1(x_1) - E(\hat{\alpha}_1(x_1))\Big)^2 dx_1\Big) \\ &= \mathcal{O}\Big(n^{-2k/(2k+1)}\Big). \end{aligned} \tag{38}$$

From (38), it follow that,

$$\int_{\mathcal{C}_1}\Big(\hat{\alpha}_1(x_1) - E(\hat{\alpha}_1(x_1))\Big)^2 dx_1 = \mathcal{O}\Big(n^{-2k/(2k+1)}\Big) \quad \text{a.s.} \tag{39}$$

By combining (37), (39) and the assumption $(H.4)$, we arrive at

$$\begin{aligned}\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}} \sup_{x_1\in I_1} T_2(x_1) &= \frac{\sup_{x_1\in I_1} \pm \alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}} \mp \int_{\mathbb{R}} \frac{\alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}} q_1(x_1)dx_1 \\ &= \frac{\sup_{x_1\in I_1} \pm \alpha_n\Big(g_n^{x_1}\Big)}{\sqrt{2h_{1,n}|\log h_{1,n}|}} + o(1) \quad \text{a.s.} \end{aligned} \tag{40}$$

Finally, note that (33), (35), (40) and Lemma 1 are sufficient to finish the proof of Theorem 1.

4.3 Proof of Theorem 2

Observe that,

$$\begin{aligned}&\sqrt{\frac{nh_{1,n}}{2|\log(h_{1,n})|}} \sup_{\mathbf{x}\in I} \pm\{\widehat{m}_{add}(\mathbf{x}) - m_{add}(\mathbf{x})\} - \sum_{l=1}^{d}\sigma_l, \\ &\quad = \sum_{l=1}^{d}\Big\{\sqrt{\frac{nh_{1,n}}{2|\log h_{1,n}|}} \sup_{x_l\in I_l} \pm\{\widehat{\eta}_l(x_l) - \eta_l(x_l)\} - \sigma_l\Big\} \\ &\qquad + \sqrt{\frac{nh_{1,n}}{2|\log(h_{1,n})|}} \int_{\mathbb{R}^d}\{\widehat{m}_n(\mathbf{x}) - m(\mathbf{x})\}q(\mathbf{x})d\mathbf{x}. \end{aligned} \tag{41}$$

By proceeding exactly as we did along the proof of (35) and (37), we arrive at, under the assumption $(H.3)$ and $(H.4)$,

$$\sqrt{\frac{nh_{1,n}}{2|\log(h_{1,n})|}}\int_{\mathbb{R}^d}\{\widehat{m}_n(\mathbf{x})-m(\mathbf{x})\}q(\mathbf{x})d\mathbf{x}=o(1)\ \text{ a.s..} \tag{42}$$

By combining Theorem 1 and the statement (42), we conclude the result of Theorem 2.□

5 Appendix

Here we gather together some basic Facts, that we need for the proofs. See, for instance Einmahl and Mason (2000) and Einmahl and Mason (2005).

Let $\mathcal{G}$ be a pointwise measurable class of functions satisfying the conditions $(\mathcal{C})$, whenever there exists a all $x \in \Xi$,

$$\mathbb{G}(\mathbf{X}) \geq \sup_{g\in\mathcal{G}} |g(x)|$$

and for some $0 < \nu, C_0 < \infty$,

$$N(\epsilon,\mathcal{G}) < C_0\epsilon^{-\nu}, 0 < \epsilon < 1,$$

with

$$N(\epsilon,\mathcal{G}) = \sup_{Q} N(\epsilon\sqrt{Q(\mathbb{G}^2)},\mathcal{G},d_Q),$$

where the supremum is taken over all probability measures Q on (Ξ, A) for which $0 < Q(\mathbb{G}^2) < \infty$ and d_Q is the $L_2(Q)$-metric. As usual $N(\epsilon,\mathcal{G},d)$ is the minimal number of balls $\{g : d(g,h) < \epsilon\}$ of d-radius ϵ needed to cover $\mathcal{G}$.

Fact 1. Let $\epsilon_1, ..., \epsilon_n$ be a sequence of a random variables independent of the random vectors $X_1, ..., X_n$. The following inequality is due to Talagrand (1994).

Let $0 < M < \infty$ be a constant, such that

$$||g|| \leq M, \quad \text{for all} \quad g \in \mathcal{G}. \tag{43}$$

Then for all $t > 0$, we have suitable finite constants $A_1, A_2 > 0$, such that

$$\mathbf{P}\Big\{||n^{1/2}\alpha_n||_{\mathcal{G}} \geq A_1\Big(E||\sum_{i=1}^{n}\epsilon_i g(\mathbf{X}_i)||_{\mathcal{G}}+t\Big)\Big\} \leq 2\Big[\exp\Big(-\frac{A_2t^2}{n\sigma^2_{\mathcal{G}}}\Big)+\exp\Big(-\frac{A_2t}{M}\Big)\Big],$$

where $\sigma^2_{\mathcal{G}} = \sup_{g\in\mathcal{G}} Var(g(\mathbf{X}))$.

Fact 2. (Einmahl and Mason (2000)) Let $\mathcal{G}$ be a pointwise measurable class of bounded functions such that for some constants ν, $C > 1$, $0 < \sigma \leq \beta$ and let $\mathbb{G}$ denote a function fulfilling the above assumptions. Assume in addition that the following four conditions hold

A.1 $E(\mathbb{G}^2(\mathbf{X})) \leq \beta^2$.
A.2 $N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu}$, $0 < \epsilon < 1$.
A.3 $\sup_{g\in\mathcal{G}} E[g^2(\mathbf{X})] \leq \sigma^2$.
A.4 $\sup_{g\in\mathcal{G}} ||g|| \leq \frac{1}{4}\sqrt{\frac{n\sigma^2}{\nu\log(C_1\beta/\sigma))}}$.

We have for a universal constant $A_3 > 0$,

$$E||\sum_{i=1}^{n} \epsilon_i g(\mathbf{X}_i)||_{\mathcal{G}} \leq A_3\sqrt{n\sigma^2\nu\log(C_1\beta/\sigma)}.$$

Fact 3. (*Bernstein's inequality*) . Let $Z_1, ..., Z_n$ be independent random variables with mean 0 and identical variance $0 < \sigma^2 := Var(Z_r) < \infty$, $r = 1, ..., n$. Assume further that for some $M > 0, |Z_r| < M, r = 1, ..., n$. Then for all $t > 0$

$$\mathbf{P}(Z_1 + ... + Z_n > t\sqrt{n}) \leq \exp\Big(- \frac{t^2}{2\sigma^2 + 2/3Mn^{-1/2}t}\Big).$$

Fact 4. Let $\eta_1, ..., \eta_n$ be independent mean zero random variables with $s_n^2 = \sum_{i=1}^n E\eta_i^2 > 0$ and $P\{|\eta_i| \leq d_i\} = 1$, where $0 < d_i \uparrow$, $1 \leq i \leq n$. if $\lim_{n\to\infty} d_n x_n/s_n = 0$, where $x_n > x_0 > 0$, then for every $\gamma \in (0, 1)$, there is a $C_\gamma \in (0, 1/2)$, such that for all large n

$$\mathbb{P}\Big\{\sum_{i=1}^{n} \eta_i \geq (1-\gamma)^2 s_n x_n\Big\} > C_\gamma \exp\Big(- x_n^2(1-\gamma)(1-\gamma^2)/2\Big).$$

References

Ango-Nze, P. and Rios, R. (2000). Density estimation in L^∞ norm for mixing processes. *J. Statist. Plann. Inference*, **83**(1), 75–90.

Camlong-Viot, C., Sarda, P., and Vieu, P. (2000). Additive time series: the kernel integration method. *Math. Methods Statist.*, **9**(4), 358–375.

Debbarh, M. (2004). Loi du logarithme itéré pour les composantes du modèle additif de régression. *C. R. Math. Acad. Sci. Paris*, **339**(10), 717–720.

Deheuvels, P. and Mason, D. (2004). General confidence bounds for nonparametric functional estimators. *Stat. Inf. for Stoch. Proc.*, **7**, 225–277.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, **13**(1), 1–37.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type functions estiamtors. *Ann. Probab.*

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. I. H. poincaré* **38**(6), 907–921.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.*, **1**(3), 297–318. With discussion.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, **82**(1), 93–100.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, **10**(2), 233–253.

Nolan, D. and Pollard (1987). U-processes: rates of convergence. *Ann. Statist*, **15**, 780–799.

Sperlich, S., Linton, O.B., and Hardle, W. (1999). Integration and backfitting methods in additive models: Finite sample properties and comparison. *Test*, **8**, 419–458.

Sperlich, S., Tjøstheim, D., and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econometric Theory*, **18**(2), 197–251.

Stone, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, (10), 1040–1053.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, **13**(2), 689–705.

Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, **22**(1), 28–76.

Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.*, **89**(428), 1398–1409.