MINIMAX REGRESSION ESTIMATION FOR POISSON COPROCESS[*]

BENOÎT CADRE[1], NICOLAS KLUTCHNIKOFF[2] AND GASPAR MASSIOT[2]

Abstract. For a Poisson point process X, Itô's famous chaos expansion implies that every square integrable regression function r with covariate X can be decomposed as a sum of multiple stochastic integrals called chaos. In this paper, we consider the case where r can be decomposed as a sum of δ chaos. In the spirit of Cadre and Truquet [*ESAIM: PS* **19** (2015) 251–267], we introduce a semiparametric estimate of r based on i.i.d. copies of the data. We investigate the asymptotic minimax properties of our estimator when δ is known. We also propose an adaptive procedure when δ is unknown.

1. INTRODUCTION

1.1. Regression estimation

Regression estimation is a central problem in statistics. It is widely used and studied in the litterature. Among all the methods explored to deal with the regression problem, nonparametric statistics have been widely investigated (see the monographies by Tsybakov [14] for a full introduction to nonparametric estimation and Györfi *et al.* [5] for a clear account on nonparametric regression). A more recent challenge regarding this statistical problem is the regression onto a functional covariate (see the books by Ramsay and Silverman [13] and Horváth and Kokozska [6] for more precision on functional data analysis). Although very challenging, the functional regression problem in the minimax setting has little coverage up to our knowledge. In the kernel estimation setting, Mas [11] studied the small ball probabilities over some Hilbert spaces to derive minimax lower bounds at fixed points. More recently, Chagny and Roche [4] derived minimax lower bounds at fixed points for adaptive nonparametric estimation of the regression under some Wiener measure domination assumptions on the small ball probabilities. Based on the k-nearest neighbor approach, Biau, Cérou and Guyader [1] used compact embedding theory to get bounds on the minimax risk. See also the references therein for a more complete overview.

1.2. Minimax regression for Poisson coprocess

In this paper, we focus on a regression problem for which the covariate is a Poisson point process. In the spirit of Cadre and Truquet [3], we use a method based on the chaotic decomposition of Poisson functionals. Let X be a Poisson point process on a compact domain $\mathbb{X} \subset \mathbb{R}^d$ equipped with its Borel σ-algebra $\mathcal{X}$. Letting δ_x the Dirac measure on $x \in \mathbb{X}$, the state space is identified to $\mathcal{S} = \{s = \sum_{i=1}^m \delta_{x_i} : m \in \mathbb{N}^*, x_i \in \mathbb{X}\}$ equipped with the smallest σ-algebra making the mappings $s \mapsto s(B)$ measurable for all Borel set B in $\mathcal{X}$. We denote by $\mathbb{P}_X$ the distribution of X whereas $\mathbb{L}^2(\mathbb{P}_X)$ denotes the space of all measurable functions $g : \mathcal{S} \to \mathbb{R}$ such that

$$\|g\|_{\mathbb{L}^2(\mathbb{P}_X)}^2 = \mathbb{E}g(X)^2 < +\infty.$$

Let $\mathbb{P}$ be a distribution on $\mathcal{S} \times \mathbb{R}$ and (X, Y) with law $\mathbb{P}$. Provided $\mathbb{E}|Y| < +\infty$ where $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$, we consider the regression function $r : \mathcal{S} \to \mathbb{R}$ defined by $r(s) = \mathbb{E}(Y \mid X = s)$.

We assume that r belongs to $\mathbb{L}^2(\mathbb{P}_X)$ and we aim at estimating r on the basis of an i.i.d. sample randomly drawn from the distribution $\mathbb{P}$ of (X, Y). In this context, any measurable map $\tilde{r} : (\mathcal{S} \times \mathbb{R})^n \to \mathbb{L}^2(\mathbb{P}_X)$ is an estimator, the accuracy of which is measured by the risk

$$R_n(\tilde{r}, r) = \mathbb{E}^n \|\tilde{r} - r\|_{\mathbb{L}^2(\mathbb{P}_X)}^2,$$

where $\mathbb{E}^n$ denotes the expectation with respect to the distribution $\mathbb{P}^{\otimes n}$. Following the minimax approach, we define the maximal risk of $\tilde{r}$ over a class $\mathcal{P}$ of distributions for the random pair (X, Y) by

$$R_n(\tilde{r}, \mathcal{P}) = \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r).$$

We are interested in finding an estimator $\hat{r}$ such that

$$R_n(\hat{r}, \mathcal{P}) \asymp \inf_{\tilde{r}} R_n(\tilde{r}, \mathcal{P}),$$

where the infimum is taken over all possible estimates of r and $u_n \asymp v_n$ stands for $0 < \liminf_n u_n v_n^{-1} \leq \limsup_n u_n v_n^{-1} < +\infty$. Such an estimate is called asymptotically minimax over $\mathcal{P}$.

1.3. Chaotic decomposition in the Poisson space

Roughly, Itô's famous chaos expansion (see Itô [8] and Nualart and Vives [12] for technical details) says that every square integrable and $\sigma(X)$-measurable random variable can be decomposed as a sum of multiple stochastic integrals, called chaos. To be more precise, we now recall some basic facts about chaos decomposition in the Poisson space. Let μ be the mean measure of the Poisson point Process X, defined by $\mu(A) = \mathbb{E}X(A)$ for $A \in \mathcal{X}$, whenever $X(A)$ is the number of points of X lying in A. Fix $k \geq 1$. Provided $g \in \mathbb{L}^2(\mu^{\otimes k})$, we can define the kth chaos $I_k(g)$ associated with g, namely

$$I_k(g) = \int_{\Delta_k} g \, \mathrm{d}(X - \mu)^{\otimes k}, \tag{1.1}$$

where $\Delta_k = \{x \in \mathbb{X}^k : x_i \neq x_j \text{ for all } i \neq j\}$. In Nualart and Vives [12], it is proved that every square integrable $\sigma(X)$-measurable random variable can be decomposed as an infinite sum of chaos. Applied to our regression problem, this statement writes as

$$r(X) = \mathbb{E}Y + \sum_{k \geq 1} \frac{1}{k!} I_k(f_k), \tag{1.2}$$

where equality holds in $\mathbb{L}^2(\mathbb{P}_X)$, provided $\mathbb{E}Y^2 < \infty$. In the above formula, each f_k is an element of $\mathbb{L}^2_{\mathrm{sym}}(\mu^{\otimes k})$ –the subset of symmetric functions in $\mathbb{L}^2(\mu^{\otimes k})$–, and the decomposition is defined in a unique way.

1.4. Organization of the paper

In this paper we introduce a new estimator of the regression function r based on independent copies of (X, Y) and we study its minimax properties. Section 2 is devoted to the definition of a semiparametric model *i.e.* the construction of the family $\mathcal{P}$ of distributions of (X, Y). In particular, we assume that r is a sum of δ chaos. In Section 3, we provide a lower bound for the minimax risk over $\mathcal{P}$. When δ is known, we prove that our estimator achieves this bound up to a logarithmic term. Finally, in Section 4, we define an adaptive procedure when δ is unknown, the risk of which is also proved to be optimal up to a logarithmic term. Last sections contain proofs.

2. MODEL

In the rest of the paper, we let $\Theta \subset \mathbb{R}^p$. For each $\theta \in \Theta$, $\varphi_\theta : \mathbb{X} \to \mathbb{R}_+$ is a Borel function. The family $\{\varphi_\theta\}_{\theta \in \Theta}$ contains the constant function $\mathbf{1}_\mathbb{X}/\lambda(X)$, and is such that there exists three positive constants $\underline{\varphi}$, $\overline{\varphi}$ and γ_1 satisfying, for all $x, y \in \mathbb{X}$ and $\theta, \theta' \in \Theta$,

$$\underline{\varphi} \leq \varphi_\theta(x) \leq \overline{\varphi}, \tag{2.1}$$

$$|\varphi_\theta(x) - \varphi_{\theta'}(x)| \leq \overline{\varphi}|\theta - \theta'|, \tag{2.2}$$

$$|\varphi_\theta(x) - \varphi_\theta(y)| \leq \gamma_1 |x - y|, \tag{2.3}$$

where, here and in the following, $|\cdot|$ stands for the euclidean norm.

Let (X, Y) be a pair of random variables taking values in $\mathcal{S} \times \mathbb{R}$ with distribution $\mathbb{P}$, where $\mathcal{S}$ is the Poisson space over the compact domain $\mathbb{X} \subset \mathbb{R}^d$. Here, X is a Poisson point process on $\mathbb{X}$ with intensity φ_θ, *i.e.* for all set $A \in \mathcal{X}$:

$$\mathbb{E}X(A) = \int_A \varphi_\theta \mathrm{d}\lambda, \tag{2.4}$$

where λ is the Lebesgue measure and $\mathbb{E}$ is the expectation with respect to $\mathbb{P}$. In other words, the mean measure of X, say μ, has a Radon–Nikodỳm derivative φ_θ with respect to λ. We assume that for all $l \geq 1$, there exists an estimator

$$\tilde{\theta}_l : \left(\mathcal{S} \times \mathbb{R}\right)^l \to \Theta,$$

such that,

$$\mathbb{E}^l |\tilde{\theta}_l - \theta|^2 \leq \frac{\kappa}{l+1}, \tag{2.5}$$

where $\kappa > 0$ is an absolute constant that does not depend on l and $\mathbb{E}^l$ is the expectation with respect to $\mathbb{P}^{\otimes l}$. As shown in Birgé ([2], Prop. 3.1), the above property is satisfied by a wide class of models, provided $\tilde{\theta}_l$ is a maximum likelihood estimate.

Moreover, the real-valued random variable Y satisfies, for some $u, M > 0$, the exponential moment condition:

$$\mathbb{E}Y^2 \mathrm{e}^{u|Y|} \leq M, \tag{2.6}$$

As seen in (1.2), the regression function $r(s) = \mathbb{E}(Y \mid X = s)$ has a chaotic decomposition. In our model, we consider the case of a finite chaotic decomposition, *i.e.* there exists a strictly positive integer δ and $f_1 \in \mathbb{L}^2_{\mathrm{sym}}(\mu), \dots, f_\delta \in \mathbb{L}^2_{\mathrm{sym}}(\mu^{\otimes \delta})$ such that

$$r(X) = \mathbb{E}Y + \sum_{k=1}^{\delta} \frac{1}{k!} I_k(f_k), \tag{2.7}$$

where the $I_k(f_k)$'s are defined in (1.1). The coefficients f_k's of the chaos lie in a nonparametric family, for which there exists two strictly positive constants γ_2 and $\bar{f}$ such that for all $k = 1, \dots, \delta$, and $x, y \in \mathbb{X}^k$

$$|f_k(x) - f_k(y)| \leq \gamma_2 |x - y|, \tag{2.8}$$

$$|f_k(x)| \leq \bar{f}. \tag{2.9}$$

Remark 2.1 (On the finiteness of the number of chaos).

(1) Finiteness of the number of chaos roughly implies that the regression function $r(X)$ is unbounded. Indeed, consider for simplicity the case where $r(X)$ is decomposed onto only one chaos, *i.e.*

$$r(X) = \int_{\mathbb{X}} f \mathrm{d}(X - \lambda).$$

Here X is a simple Poisson process on the domain $\mathbb{X} \subset \mathbb{R}$ with unit intensity and f is any λ-integrable function on $\mathbb{X}$. Observe that, if $f \geq a > 0$, then

$$r(X) \geq aX(\mathbb{X}) - \int_{\mathbb{X}} f \mathrm{d}\lambda.$$

Consequently, $r(X)$ is unbounded. The same tendency may be expected whatever the number of chaos.

(2) Finiteness of the chaotic decomposition relies on the distribution of (X, Y) *via* the Malliavin calculus. Indeed, as proved in Proposition 4.1 by Last and Penrose [10], the decomposition in δ chaos of $r(X)$ holds if and only if the $(\delta + 1)$th Malliavin derivative of r is null.

In the rest of the paper, the constants $\underline{\varphi}, \overline{\varphi}, \gamma_1, u, M, \delta, \gamma_2, \overline{f}$ and κ will be fixed, and we shall denote by $\mathcal{P}$ the set of distributions $\mathbb{P}$ of (X, Y) such that the assumptions $(2.1)-(2.9)$ are satisfied. In this setting, θ implicitly denotes the true value of the parameter, that is φ_θ is the intensity of X (with mean measure μ).

3. MINIMAX PROPERTIES FOR KNOWN δ

3.1. Chaos estimator

Our task is to construct an estimate of the regression function which achieves fast rates over $\mathcal{P}$. Let $\mathbb{P} \in \mathcal{P}$ and $(X, Y) \sim \mathbb{P}$ where X has mean measure $\mu = \varphi_\theta \cdot \lambda$.

First recall some basic facts about chaos decomposition in the Poisson space. If $g \in \mathbb{L}^2(\mu^{\otimes k})$ and $h \in \mathbb{L}^2(\mu^{\otimes l})$ for $k, l \geq 1$, we have the so-called Itô Isometry Formula:

$$\mathbb{E}I_k(g)I_l(f) = k! \int_{\mathbb{X}^k} \overline{g}\overline{h}\mathrm{d}\mu^{\otimes k}\mathbf{1}_{\{k=l\}} \text{ and } \mathbb{E}I_k(g) = 0, \tag{3.1}$$

where $\overline{g}$ and $\overline{h}$ are the symmetrizations of g and h, that is, for all $(x_1, \ldots, x_k) \in \mathbb{X}^k$:

$$\overline{g}(x_1, \ldots, x_k) = \frac{1}{k!} \sum_{\sigma} g(x_{\sigma(1)}, \ldots, x_{\sigma(k)}), \tag{3.2}$$

the sum being taken over all permutations $\sigma = \big(\sigma(1), \ldots, \sigma(k)\big)$ of $\{1, \ldots, k\}$, and similarly for $\overline{h}$.

Now let $\overline{W}$ be a strictly positive constant and W be a density on $\mathbb{X}$ such that $\sup_{\mathbb{X}} W \leq \overline{W}$. Furthermore, let $h_k = h_k(n) > 0$ a bandwidth to be tuned later on and denote

$$W_{h_k}(\cdot) = \frac{1}{h_k^d} W\left(\frac{\cdot}{h_k}\right).$$

One may easily deduce from relations (1.2) and (3.1) that

$$\mathbb{E}Y I_k\big(W_{h_k}^{\otimes k}(x - \cdot)\big) = \int_{\mathbb{X}^k} f_k W_{h_k}^{\otimes k}(x - \cdot)\varphi_\theta^{\otimes k}\mathrm{d}\lambda^{\otimes k},$$

where, here and in the following, for any real-valued function g defined on $\mathbb{X}$, the notation $g^{\otimes k}$ denotes the real-valued function on $\mathbb{X}^k$ such that

$$g^{\otimes k}(x) = \prod_{i=1}^{k} g(x_i), \quad x = (x_1, \ldots, x_k) \in \mathbb{X}^k.$$

Thus, under the smoothness assumptions (2.1) on φ_θ and (2.9) on f_k, the right-hand side converges to $f_k(x)\varphi_\theta^{\otimes k}(x)$, provided $h_k \to 0$.

Now let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. with distribution $\mathbb{P}$. Based on this observation, a semiparametric estimate denoted $\hat{I}_{k,h_k}(X)$ of the kth chaos $I_k(f_k)$ of (1.1) may be defined as follows:

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} \frac{W_{h_k}^{\otimes k}(x-y)}{\varphi_{\hat{\theta}_i}^{\otimes k}(x)} \left(X_i - \varphi_{\hat{\theta}_i} \cdot \lambda\right)^{\otimes k}(\mathrm{d}y)\left(X - \varphi_{\hat{\theta}_i} \cdot \lambda\right)^{\otimes k}(\mathrm{d}x), \tag{3.3}$$

where $T_n > 0$ is a truncation parameter to be tuned later on and the $\hat{\theta}_i$'s are the leave-one-out estimates defined by $\hat{\theta}_i = \tilde{\theta}_{n-1}\left((X_j)_{j \leq n, j \neq i}\right)$ (see Sect. 2).

3.2. Results

Based on the estimate (3.3) of the kth chaos, we may define the following empirical mean type estimator of the regression function r for any strictly positive integer l

$$\hat{r}_l(X) = \overline{Y}_n + \sum_{k=1}^{l} \frac{1}{k!} \hat{I}_{k,h_k}(X), \tag{3.4}$$

where $\overline{Y}_n$ is the empirical mean of $Y_1, \ldots, Y_n$.

In this subsection, we study the performance of the estimate $\hat{r}_\delta$ of the regression function from a minimax point of view when the number of chaos δ is known.

Theorem 3.1. *Let $\varepsilon > 0$ and set $T_n = (\ln n)^{1+\varepsilon}$ and $h_k = (T_n^2 n^{-1})^{1/(2+dk)}$. Then,*

$$\limsup_{n \to +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}}\right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}} R_n\left(\hat{r}_\delta, r\right) < \infty.$$

Remark 3.2. Thus, the optimal rate of convergence over $\mathcal{P}$ is upper bounded by $\left((\ln n)^{2+2\varepsilon} n^{-1}\right)^{2/(2+d\delta)}$. Here it is noticeable that, up to a logarithmic factor, we recover the optimal rate $n^{-2/(2+d\delta)}$ corresponding to the $d\delta$-dimensional regression with a Lipschitz regression function (see, *e.g.*, Thm. 1 in Kohler *et al.* [9]).

In our next result, we provide a lower bound for the optimal rate of convergence over $\mathcal{P}$ in order to assess the tightness of the upper bound obtained in Theorem 3.1.

Theorem 3.3. *We have,*

$$\liminf_{n \to +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P} \in \mathcal{P}} R_n(\tilde{r}, r) > 0,$$

where the infimum is taken over all estimates $\tilde{r}$.

Remark 3.4. Theorem 3.3 indicates that the optimal rate of convergence over $\mathcal{P}$ is lower bounded by $n^{-2/(2+d\delta)}$ which, up to a logarithmic factor, corresponds to the upper bound found in Theorem 3.1. As a conclusion, up to a logarithmic factor, the estimate $\hat{r}_\delta$ is asymptotically minimax on $\mathcal{P}$.

4. Adaptive properties for unknown δ

We now consider the case of an unknown number of chaos δ. For $m > 0$, we set

$$\mathcal{P}(m) = \{\mathbb{P} \in \mathcal{P} : \|f_k\| \geq m; \ k \in 1, \ldots, \delta\},$$

where $\|\cdot\|$ stands for the $\mathbb{L}^2$-norm relatively to the Lebesgue measure. Thus, whenever $\mathbb{P} \in \mathcal{P}(m)$,

$$\delta = \min(k : \|f_k\| = 0) - 1.$$

Based on this observation, a natural estimate of δ may be obtained as follows. Let we assume that the dataset is of size $2n$, and let $(X_1, Y_1), \ldots, (X_{2n}, Y_{2n})$ be i.i.d. with distribution $\mathbb{P} \in \mathcal{P}(m)$. For $k \in 1, \ldots, \delta$, we introduce the empirical counterpart of $\varphi_\theta f_k$ defined by

$$\hat{g}_k(x) = \frac{1}{n} \sum_{i=n+1}^{2n} Y_i \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y)(X_i - \varphi_{\hat{\theta}} \cdot \lambda)^{\otimes k}(\mathrm{d}y), \tag{4.1}$$

where $\hat{\theta} = \tilde{\theta}_n(X_{n+1}, \ldots, X_{2n})$ is defined in Sect. 2), and $b_k = b_k(n)$ is a bandwidth to be tuned later. The estimator $\hat{\delta}$ of δ is then defined by

$$\hat{\delta} = \min(k : \|\hat{g}_k\| \leq \rho_k) - 1, \tag{4.2}$$

where $\rho_k = \rho_k(n)$ is a vanishing sequence of positive numbers that we choose later on. We may now define the adaptive estimator $\hat{r}$ of r by

$$\hat{r} = \hat{r}_{\hat{\delta}},$$

where $\hat{r}_l$ is defined in (3.4) for all strictly positive integer l.

Theorem 4.1. *Let $\varepsilon > d\delta \geq 2$, $\alpha, \beta > 0$ such that $\alpha + \beta < 1$ and $2\alpha + \beta > 1/(2 + d\delta)$, and set $T_n = (\ln n)^{1+\varepsilon}$. Then, if we take for all integer k,*

$$h_k = (T_n^2 n^{-1})^{1/(2+dk)}, \ \rho_k = ((2k)!)^2 n^{(\alpha+\beta-1)/2} \ and \ b_k = n^{-\beta/(2dk)},$$

we obtain, for all $m > 0$,

$$\limsup_{n \to +\infty} \left(\frac{n}{(\ln n)^{2+2\varepsilon}} \right)^{2/(2+d\delta)} \sup_{\mathbb{P} \in \mathcal{P}(m)} R_n(\hat{r}, r) < +\infty.$$

Remark 4.2. Here it is noticeable that despite the estimation of the number of chaos δ and up to a logarithmic factor, we recover the optimal rate $n^{-2/(2+d\delta)}$ of Theorems 3.1 and 3.3.

5. Proof of Theorem 3.1

In this section, we assume without loss of generality that the constants $\overline{\varphi}$, γ_1, γ_2, $\bar{f}$, $\lambda(\mathbb{X})$ and $\overline{W}$ are greater than 1 and that $\underline{\varphi}$ is smaller than 1. Moreover, C denotes a positive number that only depends on the parameters of the model, *i.e.* $u, \underline{\varphi}, \overline{\varphi}, \gamma_1, \gamma_2, \bar{f}, \delta, \theta, \kappa, \lambda(\mathbb{X}), M$ and W, and whose value may change from line to line.

We let $\mathbb{P} \in \mathcal{P}$ and, for simplicity, we may denote $\mathbb{E} = \mathbb{E}^n$ and var stands for the variance relatively to $\mathbb{P}^{\otimes n}$. Finally, let $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d. with distribution $\mathbb{P}$.

5.1. Technical results

Let $k \geq 1$ be fixed and denote for all $x, y \in \mathbb{X}$ and $i = 1, \ldots, n$:

$$g_i(x, y) = \frac{W_{h_k}(x - y)}{\varphi_{\hat{\theta}_i}(x)} \text{ and } g(x, y) = \frac{W_{h_k}(x - y)}{\varphi_{\theta}(x)}. \tag{5.1}$$

We also let

$$\mathrm{d}\hat{X}_i = \mathrm{d}X_i - \varphi_{\hat{\theta}_i}\mathrm{d}\lambda, \ \mathrm{d}\hat{X}'_i = \mathrm{d}X - \varphi_{\hat{\theta}_i}\mathrm{d}\lambda, \tag{5.2}$$

$$\mathrm{d}\tilde{X}_i = \mathrm{d}X_i - \varphi_{\theta}\mathrm{d}\lambda \text{ and } \mathrm{d}\tilde{X} = \mathrm{d}X - \varphi_{\theta}\mathrm{d}\lambda. \tag{5.3}$$

With this respect, we have (see (3.3)):

$$\hat{I}_{k,h_k}(X) = \frac{1}{n}\sum_{i=1}^{n} Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k^2} g_i^{\otimes k}(x, y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\hat{X}_i'^{\otimes k}(\mathrm{d}x).$$

Furthermore, denote for all $x \in \mathbb{X}^k$:

$$Z_{i,k}(x) = Y_i \mathbf{1}_{|Y_i| \leq T_n} \int_{\Delta_k} g^{\otimes k}(x, y)\tilde{X}_i^{\otimes k}(\mathrm{d}y). \tag{5.4}$$

Lemma 5.1. *Let $i = 1, \ldots, n$ and $k \leq \delta$ be fixed. Then for all $x \in \mathbb{X}^k$:*

$$\mathrm{var}\big(Z_{i,k}(x)\big) \leq T_n^2 \frac{k! C^k}{h_k^{dk}}, \quad \text{and} \quad |\mathbb{E}Z_{i,k}(x) - f_k(x)| \leq C^k \left(\phi\frac{k!}{h_k^{dk}}\right)^{1/2} + C^k h_k,$$

where $\phi = \mathbb{E}Y^2\mathbf{1}_{|Y|>T_n}$.

Proof. On the one hand, by the isometry formula (3.1) over the set $\mathcal{P}$,

$$\mathrm{var}\big(Z_{i,k}(x)\big) \leq T_n^2 \mathbb{E}\left(\int_{\Delta_k} g^{\otimes k}(x, y)\tilde{X}^{\otimes k}(\mathrm{d}y)\right)^2$$

$$\leq T_n^2 k! \int_{\mathbb{X}^k} \overline{g^{\otimes k}}^2(x, y)\varphi_{\theta}^{\otimes k}(y)\mathrm{d}y$$

$$\leq T_n^2 \frac{\overline{W}^k \overline{\varphi}^k}{\underline{\varphi}^{2k}} \frac{k!}{h_k^{dk}},$$

where $\overline{g^{\otimes k}}(x, \cdot)$ is the symmetrization –see (3.2)– of the function $g^{\otimes k}(x, \cdot)$ defined in (5.1). On the other hand, denote

$$\tilde{Z}_{i,k}(x) = Y_i \int_{\Delta_k} g^{\otimes k}(x, y)\tilde{X}_i^{\otimes k}(\mathrm{d}y),$$

then by the isometry formula (3.1):

$$\mathbb{E}\tilde{Z}_{i,k}(x) = \mathbb{E}r(X)\int_{\Delta_k} g^{\otimes k}(x, y)\tilde{X}^{\otimes k}(\mathrm{d}y) = \int_{\mathbb{X}^k} f_k(y)\overline{g^{\otimes k}}(x, y)\varphi_{\theta}^{\otimes k}(y)\mathrm{d}y$$

$$= \frac{1}{\varphi_{\theta}^{\otimes k}(x)}\int_{\mathbb{X}^k} f_k(x - h_k z)\overline{W^{\otimes k}}(z)\varphi_{\theta}^{\otimes k}(x - h_k z)\mathrm{d}z.$$

Furthermore, by assumptions (2.1), (2.3), (2.8) and (2.9) on the model, we have for all $x, y \in \mathbb{X}^k$:

$$
\begin{aligned}
|f_k(x)\varphi_\theta^{\otimes k}(x) - f_k(y)\varphi_\theta^{\otimes k}(y)| &\leq \overline{\varphi}^k |f_k(x) - f_k(y)| + \bar{f}|\varphi_\theta^{\otimes k}(x) - \varphi_\theta^{\otimes k}(y)| \\
&\leq \left(\overline{\varphi}^k \gamma_2 + k\bar{f}\overline{\varphi}^{k-1}\gamma_1\right)|x - y| \\
&\leq 2k\bar{f}\overline{\varphi}^k \gamma_2\gamma_1|x - y|.
\end{aligned}
$$

Hence, letting $\omega_k = \int_{\mathbb{X}^k} |z| W^{\otimes k}(z)\mathrm{d}z$,

$$
\begin{aligned}
|\mathbb{E}Z_{i,k}(x) - f_k(x)| &\leq |\mathbb{E}(Z_{i,k}(x) - \tilde{Z}_{i,k}(x))| + |\mathbb{E}\tilde{Z}_{i,k}(x) - f_k(x)| \\
&\leq \left|\mathbb{E}Y\mathbf{1}_{|Y|>T_n}\int_{\Delta_k} g^{\otimes k}(x,y)\tilde{X}^{\otimes k}(\mathrm{d}y)\right| + |\mathbb{E}\tilde{Z}_{i,k}(x) - f_k(x)| \\
&\leq \phi^{1/2}\left(\mathbb{E}\left(\int_{\Delta_k} g^{\otimes k}(x,y)\tilde{X}^{\otimes k}(\mathrm{d}y)\right)^2\right)^{1/2} + 2k\bar{f}\omega_k\frac{\overline{\varphi}^k \gamma_2\gamma_1}{\underline{\varphi}^k}h_k.
\end{aligned}
$$

One last application of the isometry formula (3.1) to the first term on the right-hand side of above gives the Proposition. $\qquad\square$

With the help of notations (5.1)–(5.3), define

$$
R_{1k}^i = \mathbb{E}\left(\int_{\Delta_k^2} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\left[\hat{X}_i'^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)\right]\right)^2, \tag{5.5}
$$

$$
R_{2k}^i = \mathbb{E}\left(\int_{\Delta_k^2} g_i^{\otimes k}(x,y)\left[\hat{X}_i^{\otimes k}(\mathrm{d}y) - \tilde{X}_i^{\otimes k}(\mathrm{d}y)\right]\tilde{X}^{\otimes k}(\mathrm{d}x)\right)^2. \tag{5.6}
$$

Lemma 5.2. *Let $i = 1, \ldots, n$ and $k \leq \delta$ be fixed. Then, for $j = 1$ or 2:*

$$
R_{jk}^i \leq C^k \frac{(k!)^2}{nh_k^{dk}}.
$$

Proof. The proofs for the bounds for R_{1k}^i and R_{2k}^i being similar, we only prove the one for R_{1k}^i. We have

$$
R_{1k}^i = \mathbb{E}\mathbb{E}\left[\left(\int_{\Delta_k^2} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\left[\hat{X}_i'^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)\right]\right)^2\Big|(X_l)_{l\leq n}\right],
$$

Using the independence of X and $(X_l)_{l\leq n}$, we can apply Lemma 4.2 from Cadre and Truquet [3], which entails

$$
R_{1k}^i \leq \sum_{j=0}^{k-1} j!\binom{k}{j}^2 \overline{\varphi}^j \int_{\mathbb{X}^k} \mathbb{E}\left[V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\right)^2\right]\mathrm{d}x, \tag{5.7}
$$

where $V_i = \|\varphi_{\hat{\theta}_i} - \varphi_\theta\|^2$. Now let $x \in \mathbb{X}^k$ and $j = 0, \ldots, k-1$ be fixed. We have

$$
\begin{aligned}
\mathbb{E}V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\right)^2 &\leq 2\mathbb{E}V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\left(\hat{X}_i^{\otimes k}(\mathrm{d}y) - \tilde{X}_i^{\otimes k}(\mathrm{d}y)\right)\right)^2 \\
&\quad + 2\mathbb{E}V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\tilde{X}_i^{\otimes k}(\mathrm{d}y)\right)^2.
\end{aligned}
$$

We proceed to bound the two terms on the right-hand side of above. As before, we apply Lemma 4.2 from Cadre and Truquet [3], but conditionally on $(X_l)_{l\leq n, l\neq i}$. For notational simplicity, and since it does change the result anymore, we do not specify the symmetrized version of the functions when using the isometry formula. By (3.1) and assumption (2.1), we then get

$$\mathbb{E}V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\right)^2 \leq 2\sum_{l=0}^{k-1} l!\binom{k}{l}^2 \overline{\varphi}^l \mathbb{E}V_i^{2k-l-j}\int_{\mathbb{X}^k} g_i^{\otimes k}(x,y)^2\mathrm{d}y$$
$$+ 2k!\overline{\varphi}^k \mathbb{E}V_i^{k-j}\int_{\mathbb{X}^k} g_i^{\otimes k}(x,y)^2\mathrm{d}y.$$

Note that for all $m \geq 1$, by (2.2), we have

$$V_i^m \leq \left(\overline{\varphi}^2\lambda(\mathbb{X})\right)^{m-1}\left(\sup_{\mathbb{X}}|\varphi_{\hat{\theta}_i} - \varphi_\theta|\right)^2\lambda(\mathbb{X})$$
$$\leq \left(\overline{\varphi}^2\lambda(\mathbb{X})\right)^m |\hat{\theta}_i - \theta|^2,$$

and

$$\int_{\mathbb{X}^k} g_i^{\otimes k}(x,y)^2\mathrm{d}y \leq \frac{\overline{W}^k}{\underline{\varphi}^{2k}h_k^{dk}}.$$

Hence, since $l!\binom{k}{l} \leq k!$, we get with (2.5):

$$\mathbb{E}V_i^{k-j}\left(\int_{\Delta_k} g_i^{\otimes k}(x,y)\hat{X}_i^{\otimes k}(\mathrm{d}y)\right)^2 \leq 2k!\frac{\overline{W}^k}{\underline{\phi}^{2k}}\frac{\kappa}{nh_k^{dk}}\left(\overline{\phi}^2\lambda(\mathbb{X})\right)^{k-j}\left(\overline{\varphi} + \overline{\varphi}^2\lambda(\mathbb{X})\right)^k.$$

Finally, we deduce with similar arguments and inequality (5.7) that

$$R_{1k}^i \leq 2(k!)^2\frac{\overline{W}^k}{\underline{\varphi}^{2k}}\frac{\kappa}{nh_k^{dk}}\left(\overline{\varphi} + \overline{\varphi}^2\lambda(\mathbb{X})\right)^{2k}$$
$$\leq 2(k!)^2\frac{4^k\overline{W}^k\overline{\varphi}^{4k}}{\underline{\varphi}^{2k}}\frac{\kappa}{nh_k^{dk}}\lambda(\mathbb{X})^{2k},$$

because both $\overline{\varphi}$ and $\lambda(\mathbb{X})$ are greater than 1. The Lemma is proved. $\qquad\square$

Lemma 5.3. *Let $\varepsilon > 0$ be fixed and set $T_n = (\ln n)^{1+\varepsilon}$. Then, for all $k \leq \delta$:*

$$\mathbb{E}\left(\hat{I}_{k,h_k}(X) - I_k(f_k)\right)^2 \leq C^k(k!)^2\left(\frac{(\ln n)^{2+2\varepsilon}}{nh_k^{dk}} + h_k^2\right).$$

Proof. With the help of notations (5.1)−(5.3), we let

$$J_1 = \frac{1}{n}\sum_{i=1}^n Y_i\mathbf{1}_{|Y_i|\leq T_n}\int_{\Delta_k^2} g_i^{\otimes k}(x,y)\tilde{X}_i^{\otimes k}(\mathrm{d}y)\tilde{X}^{\otimes k}(\mathrm{d}x),$$
$$J_2 = \frac{1}{n}\sum_{i=1}^n Y_i\mathbf{1}_{|Y_i|\leq T_n}\int_{\Delta_k^2} g^{\otimes k}(x,y)\tilde{X}_i^{\otimes k}(\mathrm{d}y)\tilde{X}^{\otimes k}(\mathrm{d}x).$$

Then, using notations of Lemma 5.2, by Jensen's Inequality

$$\mathbb{E}\big(\hat{I}_{k,h_k}(X) - J_1\big)^2 \le 2\frac{T_n^2}{n}\sum_{i=1}^n (R_{1k}^i + R_{2k}^i),$$

Hence, by Lemma 5.2

$$\mathbb{E}\big(\hat{I}_{k,h_k}(X) - J_1\big)^2 \le T_n^2 C^k \frac{(k!)^2}{nh_k^{dk}}. \tag{5.8}$$

Moreover, sequentially conditioning on $(X_l)_{l \le n}$, then on $(X_l)_{l \le n, l \ne i}$, and using assumption (2.1), we find with two successive applications of the isometry formula (3.1) that

$$\mathbb{E}\big(J_1 - J_2\big)^2 \le (k!)^2 T_n^2 \overline{\varphi}^{2k} \mathbb{E} \int_{\mathbb{X}^{2k}} \left(\overline{g_1^{\otimes k}(x,y) - g^{\otimes k}(x,y)} \right)^2 \mathrm{d}x\mathrm{d}y.$$

Now let $x, y \in \mathbb{X}^k$ be fixed. We have

$$\big|g_1^{\otimes k}(x,y) - g^{\otimes k}(x,y)\big| \le \frac{W_{h_k}^{\otimes k}(x-y)}{\varphi_\theta^{\otimes k}(x)} k \frac{\overline{\varphi}^{k-1}}{\underline{\varphi}^k} \sup_{\mathbb{X}} |\varphi_{\hat{\theta}_1} - \varphi_\theta|,$$

so that (2.2) and (2.5) give

$$\mathbb{E}\big(J_1 - J_2\big)^2 \le C^k T_n^2 \frac{(k!)^2}{nh_k^{dk}}. \tag{5.9}$$

Finally, using notation (5.4), by the isometry formula (3.1), we have

$$\begin{aligned}
\mathbb{E}\big(J_2 - I_k(f_k)\big)^2 &= \mathbb{E}\left(\int_{\Delta_k} \left(\frac{1}{n}\sum_{i=1}^n Z_{i,k}(x) - f_k(x) \right) \tilde{X}^{\otimes k}(\mathrm{d}x) \right)^2 \\
&= k! \int_{\mathbb{X}^k} \mathbb{E}\big(\frac{1}{n}\sum_{i=1}^n Z_{i,k}(x) - f_k(x)\big)^2 \varphi_\theta^{\otimes k}(x)\mathrm{d}x \\
&= k! \int_{\mathbb{X}^k} \left(\frac{1}{n}\mathrm{var}\big(Z_{1,k}(x)\big) + \big(\mathbb{E}Z_{1,k}(x) - f_k(x)\big)^2 \right) \varphi_\theta^{\otimes k}(x)\mathrm{d}x.
\end{aligned}$$

By Lemma 5.1, we thus get

$$\mathbb{E}\big(J_2 - I_k(f_k)\big)^2 \le T_n^2 \frac{(k!)^2 C^k}{nh_k^{dk}} + C^k (k!)^2 \frac{\phi^2}{h_k^{dk}} + C^k k! h_k^2.$$

Moreover, given that (2.6) gives $\phi \le \mathrm{e}^{-uT_n}\mathbb{E}Y^2\mathrm{e}^{u|Y|}$. Consequently,

$$\mathbb{E}\big(J_2 - I_k(f_k)\big)^2 \le C^k (k!)^2 \left(\frac{T_n^2}{nh_k^{dk}} + \frac{\mathrm{e}^{-2uT_n}}{h_k^{dk}} + h_k^2 \right).$$

Finally, combining inequalities (5.8), (5.9) and above, we deduce that with the choice $T_n = (\ln n)^{1+\varepsilon}$:

$$\mathbb{E}\big(\hat{I}_{k,h_k}(X) - I_k(f_k)\big)^2 \le C^k (k!)^2 \left(\frac{(\ln n)^{2+2\varepsilon}}{nh_k^{dk}} + h_k^2 \right),$$

hence the Lemma. $\qquad\qquad\square$

5.2. Proof of Theorem 3.1

According to Jensen Inequality and Lemma 5.3, we have by (3.4) and (2.7):

$$\mathbb{E}\big(\hat{r}_\delta(X) - r(X)\big)^2 = \mathbb{E}\left(\bar{Y}_n - \mathbb{E}Y + \sum_{k=1}^{\delta} \frac{1}{k!}\big(\hat{I}_{k,h_k}(X) - I_k(f_k)\big)\right)^2$$

$$\leq \frac{2\mathrm{var}(Y)}{n} + \delta \sum_{k=1}^{\delta} C^k \left(\frac{(\ln n)^{2+2\varepsilon}}{nh_k^{dk}} + h_k^2\right).$$

Setting $h_k = \big((\ln n)^{2+\varepsilon} n^{-1}\big)^{1/(2+dk)}$, we deduce that, since $\mathrm{var}(Y) \leq M$:

$$\mathbb{E}\big(\hat{r}_\delta(X) - r(X)\big)^2 \leq \frac{2M}{n} + C \left(\frac{(\ln n)^{2+2\varepsilon}}{n}\right)^{2/(2+d\delta)},$$

hence the theorem. $\qquad\square$

6. PROOF OF THEOREM 3.3

In this section we assume for simplicity that $\mathbb{X}$ contains the hypercube $\mathbb{X}_0 = [0,1]^d$.

6.1. Technical results

We introduce the set $\mathcal{F} = \mathcal{F}_\delta(\gamma_2, \bar{f})$ of functions $f : \mathbb{X}^\delta \to \mathbb{R}$ in $\mathbb{L}^2_{\mathrm{sym}}(\lambda^{\otimes\delta})$ for which conditions (2.8) and (2.9) hold, and let $\mathcal{R} = \mathcal{R}_\delta(\gamma_2, \bar{f})$ be the class of functions $r_f : \mathcal{S} \to \mathbb{R}$ with $f \in \mathcal{F}$ such that

$$r_f(\cdot) = \frac{1}{\delta!} \int_{\Delta_\delta} f \mathrm{d}(\cdot - \lambda)^{\otimes\delta}. \tag{6.1}$$

Letting P the distribution of the Poisson point process on $\mathbb{X}$ with unit intensity, we may define the following distance D on $\mathcal{R}$ by

$$D(r_{f_0}, r_{f_1}) = \|r_{f_0} - r_{f_1}\|_{\mathbb{L}^2(P)}. \tag{6.2}$$

Whenever $\mathbb{P} \in \mathcal{P}$, we can associate the regression function r. To stress the dependency on r, we now write $\mathbb{P}_r$ instead of $\mathbb{P}$. Now let $N > 0$. We define the following three conditions for any sequence of size $N+1$ of functions $r^{(0)}, \ldots, r^{(N)}$ from $\mathcal{S}$ to $\mathbb{R}$:

(R1) $r^{(j)} \in \mathcal{R}$, for $j = 0, \ldots, N$;
(R2) $D(r^{(i)}, r^{(j)}) \geq 2n^{-1/(2+d\delta)}$, for $0 \leq i < j \leq N$;
(R3) $\dfrac{1}{N} \sum_{j=1}^{N} \mathcal{K}\big(\mathbb{P}_{r^{(j)}}^{\otimes n}, \mathbb{P}_{r^{(0)}}^{\otimes n}\big) \leq \alpha \log N$ for some $0 < \alpha < 1/8$ where $\mathcal{K}$ is the Kullback-Leibler divergence (see *e.g.* Tsybakov [14]).

Lemma 6.1. *Introduce $f_0 \equiv 0, f_1, \ldots, f_N$, $N+1$ functions from $\mathbb{X}^\delta$ to $\mathbb{R}$ such that*

(F1) $f_j \in \mathcal{F}$;
(F2) $\|f_i - f_j\| \geq 2n^{-1/(2+d\delta)}$;
(F3) $\dfrac{1}{N} \sum_{i=1}^{N} \dfrac{n}{2} \|f_i\|^2 \leq \alpha \log N$ *for some* $0 < \alpha < 1/8$.

Then, the sequence of functions $r_{f_0}, \ldots, r_{f_N}$ defined by (6.1) *verify conditions* R1, R2 *and* R3.

Proof. First of all, remark that since $f_j \in \mathcal{F}$ for $j = 0, \ldots, N$, by definition of the r_{f_j}'s and the set $\mathcal{R}$, we have $r_{f_j} \in \mathcal{R}$. Hence condition R1 is satisfied by the r_{f_j}'s. Now remark that the Itô isometry (3.1) gives for any $0 \le i, j \le N$

$$D(r_{f_j}, r_{f_i}) = \|f_j - f_i\|.$$

This ensures that the r_{f_j}'s statisfy condition R2. Finally, for all $j = 0, \ldots, N$

$$
\begin{aligned}
\mathcal{K}\big(\mathbb{P}_{r_{f_j}}^{\otimes n}, \mathbb{P}_{r_{f_0}}^{\otimes n}\big) &= n\mathcal{K}(\mathbb{P}_{r_{f_j}}, \mathbb{P}_{r_{f_0}}) \\
&= n\mathbb{E}_{r_{f_0}}\left(\log \frac{\mathrm{d}\mathbb{P}_{r_{f_0}}}{\mathrm{d}\mathbb{P}_{r_{f_j}}}(X, Y)\right) \\
&= n\mathbb{E}_{r_{f_0}}\mathbb{E}_{r_{f_0}}\left(\log \frac{\mathrm{d}\mathbb{P}_{r_{f_0}}}{\mathrm{d}\mathbb{P}_{r_{f_j}}}(X, Y) \mid X\right),
\end{aligned}
$$

where $\mathbb{E}_{r_{f_0}}$ is the expectation under $\mathbb{P}_{r_{f_0}}$. Denote by p the density of the $\mathcal{N}(0, 1)$. Then, since $f_0 \equiv 0$

$$\mathbb{E}_{r_{f_0}}\left(\log \frac{\mathrm{d}\mathbb{P}_{r_{f_0}}}{\mathrm{d}\mathbb{P}_{r_{f_j}}}(X, Y) \mid X\right) = \int_{\mathbb{R}} \log\left(\frac{p(u)}{p\big(u - r_{f_j}(X)\big)}\right)p(u)\mathrm{d}u.$$

Simple calculus then give

$$\mathbb{E}_{r_{f_0}}\left(\log \frac{\mathrm{d}\mathbb{P}_{r_{f_0}}}{\mathrm{d}\mathbb{P}_{r_{f_j}}}(X, Y) \mid X\right) \le \frac{1}{2}\big(r_{f_j}(X)\big)^2.$$

Thus, by the Itô Isometry,

$$\mathcal{K}\big(\mathbb{P}_{r_{f_j}}^{\otimes n}, \mathbb{P}_{r_{f_0}}^{\otimes n}\big) \le \frac{n}{2}\|f_j\|^2,$$

hence the lemma. $\square$

6.2. Proof of Theorem 3.3

Let $\mathcal{P}_0$ be the subset of distributions $\mathbb{P}_r$ of (X, Y) in $\mathcal{P}$ for which X is a Poisson point process with unit intensity (recall that the unit function lies in $\{\varphi_\theta\}_{\theta \in \Theta}$, see Sect. 2) and such that

$$Y = r(X) + \varepsilon, \tag{6.3}$$

where $r \in \mathcal{R}$ and ε is independent from X with distribution $\mathcal{N}(0, 1)$. Since $\mathcal{P}_0 \subset \mathcal{P}$, we have

$$\inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} R_n(\tilde{r}, r) \le \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}} R_n(\tilde{r}, r).$$

As a result, in order to prove Theorem 3.3, we need only to prove that

$$\liminf_{n \to +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} R_n(\tilde{r}, r) > 0,$$

which accordingly to (6.2) may be written

$$\liminf_{n \to +\infty} n^{2/(2+d\delta)} \inf_{\tilde{r}} \sup_{\mathbb{P}_r \in \mathcal{P}_0} \mathbb{E}_r^n D^2(\tilde{r}, r) > 0, \tag{6.4}$$

where $\mathbb{E}_r^n$ denotes expectation with respect to $\mathbb{P}_r^{\otimes n}$. Then, according to Lemma 6.1 and Theorem 2.5 page 99 in the book by Tsybakov [14], in order to prove (6.4), we need only to prove the existence of a sequence of functions satisfying conditions F1, F2 and F3 defined in the previous subsection. To this end, we let $\psi \in \mathbb{L}^2_{\text{sym}}(\lambda^{\otimes \delta})$ be a nonzero function such that $\text{Supp}(\psi) = \mathbb{X}_0^\delta$ and, for all $x, y \in \mathbb{X}_0^\delta$:

$$|\psi(y) - \psi(x)| \leq \frac{\gamma_2}{2}|y - x| \text{ and } |\psi(x)| \leq \bar{f}. \tag{6.5}$$

Let $Q = \lfloor c_0 n^{d\delta/(2+d\delta)} \rfloor \geq 8$ where $c_0 > 0$ and $\lfloor \cdot \rfloor$ is the integer part and let $a_n = (1/Q)^{1/(d\delta)}$. One may easily prove that there exists $t_1, \ldots, t_Q$ in $\mathbb{X}_0^\delta$ such that the functions

$$\psi_q(\cdot) = a_n \psi\left(\frac{\cdot - t_q}{a_n}\right), \qquad \text{for} \qquad q = 1, \ldots, Q,$$

verify the following assumptions

(1) $\text{Supp}(\psi_q) \subset \mathbb{X}_0^\delta$, for $q = 1, \ldots, Q$;
(2) $\text{Supp}(\psi_q) \cap \text{Supp}(\psi_q') = \emptyset$, for $q \neq q'$;
(3) $\lambda^{\otimes \delta}(\text{Supp}(\psi_q)) = Q^{-1}$.

Now let, for all $\omega \in \{0, 1\}^Q$:

$$f_\omega(\cdot) = \sum_{q=1}^Q \omega_q \psi_q(\cdot),$$

According to the Varshamov$-$Gilbert Lemma (see the book by Tsybakov [14], Lem. 2.8 p. 104), there exists a subset $\Omega = \{\omega^{(0)}, \ldots, \omega^{(N)}\}$ of $\{0, 1\}^Q$ such that $\omega^{(0)} = (0, 0, \ldots)$, $N \geq 2^{Q/8}$, and for all $j \neq k$:

$$\sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \geq \frac{Q}{8}.$$

Now fix $0 < \alpha < 1/8$ and set

$$c_0 = (4\|\psi\|^2 \alpha^{-1})^{d\delta/(2+d\delta)}.$$

We may now prove that functions $\{f_{\omega^{(j)}} : j = 0, \ldots, N\}$ satisfy conditions F1, F2 and F3. First of all, let $j \neq k$ be fixed and remark that,

$$\|f_{\omega^{(j)}} - f_{\omega^{(k)}}\|^2 \leq \int_{\mathbb{X}_0^\delta} \left(f_{\omega^{(j)}}(x) - f_{\omega^{(k)}}(x)\right)^2 \mathrm{d}x$$

$$= \int_{\mathbb{X}_0^\delta} \left(\sum_{q=1}^Q (\omega_q^{(j)} - \omega_q^{(k)})\psi_q(x)\right)^2 \mathrm{d}x$$

$$= \sum_{q=1}^Q \int_{\text{Supp}(\psi_q)} (\omega_q^{(j)} - \omega_q^{(k)})^2 a_n^2 \psi^2\left(\frac{x - t_q}{a_n}\right) \mathrm{d}x,$$

$$= \frac{a_n^2}{Q} \sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \int_{\mathbb{X}_0^\delta} \psi^2(x) \mathrm{d}x.$$

Furthermore, by definition of the set Ω, we have

$$\frac{Q}{8} \leq \sum_{q=1}^Q |\omega_q^{(j)} - \omega_q^{(k)}| \leq Q,$$

so that

$$\frac{\|\psi\|^2}{8} a_n^2 \leq \|f_{\omega^{(j)}} - f_{\omega^{(k)}}\|^2 \leq \|\psi\|^2 a_n^2. \tag{6.6}$$

Now let $0 \leq j \leq N$. Since $\psi \in \mathbb{L}^2_{\text{sym}}(\lambda^{\otimes \delta})$ it is clear that $f_{\omega^{(j)}}$ inherits that property. Then, using the first part of assumption (6.5) on ψ and assumption (ii) on the ψ_q's, one may easily prove that $f_{\omega^{(j)}}$ is a Lipschitz function with constant γ_2. Finally, using the second part of assumption (6.5) on ψ and assumption (ii) on the ψ_q's, we get that for all $x \in \mathbb{X}^k$, $|f_{\omega^{(j)}}(x)| \leq a_n \bar{f}$ with $a_n \leq 1$ as soon as $n \geq c_0^{-(2+d\delta)/(d\delta)}$. We conclude that $f_{\omega^{(j)}} \in \mathcal{F}$ so that condition F1 is satisfied by $f_{\omega^{(j)}}$. Now, taking $k = 0$ in (6.6) gives

$$\frac{1}{N} \sum_{j=1}^{N} \frac{n}{2} \|f_{\omega^{(j)}}\|^2 \leq \frac{n\|\psi\|^2}{2} a_n^2 \leq \frac{\|\psi\|^2}{2} c_0^{-(2+d\delta)/(d\delta)} Q.$$

Since $N \geq 2^{Q/8}$ and $c_0 \geq (4\|\psi\|^2 \alpha^{-1})^{d\delta/(2+d\delta)}$, we get

$$\frac{1}{N} \sum_{j=1}^{N} \frac{n}{2} \|f_{\omega^{(j)}}\|^2 \leq 4\|\psi\|^2 c_0^{-(2+d\delta)/(d\delta)} \log N \leq \alpha \log N,$$

so that F3 is satisfied. Finally, according to (6.6), we have

$$\|f_{\omega^{(j)}} - f_{\omega^{(k)}}\| \geq \frac{\|\psi\|}{2\sqrt{2}} a_n = \frac{\|\psi\|}{2\sqrt{2}} (2c_0)^{-1/(d\delta)} n^{-1/(2+d\delta)}.$$

We can now conclude that conditions F1, F2 and F3 are satisfied so that according to Theorem 2.5 page 99 from the book by Tsybakov [14] and Lemma 6.1, (6.4) is verified. Theorem follows. □

7. Proof of Theorem 4.1

In this section, we assume without loss of generality that the constants $\overline{\varphi}$, γ_1, γ_2, $\bar{f}$, $\lambda(\mathbb{X})$ and $\overline{W}$ are greater than 1 and that $\underline{\varphi}$ and m are smaller than 1. Moreover, C denotes a positive number that only depends on the parameters of the model, $i.e.$ $m, u, \underline{\varphi}, \overline{\varphi}, \delta, \gamma_1, \gamma_2, \bar{f}, \theta, \kappa, \lambda(\mathbb{X}), M$ and W, and whose value may change from line to line.

We let $\mathbb{P} \in \mathcal{P}(m)$, and for simplicity, we denote $\mathbb{E} = \mathbb{E}^{2n}$ the expectation with respect to $\mathbb{P}^{\otimes 2n}$. Recall that the pairs $(X, Y), (X_1, Y_1), \ldots, (X_{2n}, Y_{2n})$ are i.i.d. with distribution $\mathbb{P}$.

7.1. Technical results

For $k \geq 1$, denote for all $x \in \mathbb{X}^k$:

$$g_k(x) = \frac{1}{n} \sum_{i=1}^{n} Y_i \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y)(X_i - \varphi_\theta \cdot \lambda)^{\otimes k}(\mathrm{d}y). \tag{7.1}$$

We also let

$$S_k = \frac{1}{n} \sum_{i=1}^{n} |Y_i|(X_i(\mathbb{X}) + \overline{\varphi}\lambda(\mathbb{X}))^k. \tag{7.2}$$

Lemma 7.1. *We have, for all* k:

$$\|\hat{g}_k - g_k\| \leq C^k S_k \frac{|\hat{\theta} - \theta|}{b_k^{dk/2}}.$$

Proof. Let $\hat{\mu}$ the signed measure

$$\hat{\mu}(\mathrm{d}y) = \frac{1}{n}\sum_{i=1}^{n} Y_i\left[\left(X_i - \varphi_{\hat{\theta}}\cdot\lambda\right)^{\otimes k}(\mathrm{d}y) - \left(X_i - \varphi_{\theta}\cdot\lambda\right)^{\otimes k}(\mathrm{d}y)\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i \sum_{l=0}^{k-1}\binom{k}{l}(-1)^{k-l}\prod_{j=1}^{l}X_i(\mathrm{d}y_j)\left[\prod_{j=l+1}^{k}\varphi_{\hat{\theta}}(y_j)\mathrm{d}y_j - \prod_{j=l+1}^{k}\varphi_{\theta}(y_j)\mathrm{d}y_j\right].$$

Then,

$$\|\hat{g}_k - g_k\|^2 = \int_{\mathbb{X}^k}\left(\int_{\Delta_k}W_{b_k}^{\otimes k}(x-y)\hat{\mu}(\mathrm{d}y)\right)^2\mathrm{d}x$$

$$= \int_{\Delta_k^2}\left[\int_{\mathbb{X}^k}W_{b_k}^{\otimes k}(x-y^{(1)})W_{b_k}^{\otimes k}(x-y^{(2)})\mathrm{d}x\right]\hat{\mu}(\mathrm{d}y^{(1)})\hat{\mu}(\mathrm{d}y^{(2)}).$$

But,

$$\int_{\mathbb{X}^k}W_{b_k}^{\otimes k}(x-y^{(1)})W_{b_k}^{\otimes k}(x-y^{(2)})\mathrm{d}x \leq \frac{\overline{W}^k}{b_k^{dk}}.$$

Furthermore, one gets by induction that for $l = 0,\ldots,k-1$:

$$\left|\prod_{j=l+1}^{k}\varphi_{\hat{\theta}}(y_j) - \prod_{j=l+1}^{k}\varphi_{\theta}(y_j)\right| \leq \overline{\varphi}^{k-l-1}\sum_{j=l+1}^{k}|\varphi_{\hat{\theta}}(y_j) - \varphi_{\theta}(y_j)|,$$

and by assumptions (2.1) and (2.2),

$$\int_{\mathbb{X}^{k-l}}\sum_{j=l+1}^{k}|\varphi_{\hat{\theta}}(y_j) - \varphi_{\theta}(y_j)|\mathrm{d}y_{l+1}\ldots\mathrm{d}y_k \leq (k-l)\lambda(\mathbb{X})^{k-l-1}\!\!\int_{\mathbb{X}}|\varphi_{\hat{\theta}} - \varphi_{\theta}|\mathrm{d}\lambda$$

$$\leq (k-l)\lambda(\mathbb{X})^{k-l}\overline{\varphi}|\hat{\theta} - \theta|.$$

Puting all pieces together, we get

$$\|\hat{g}_k - g_k\|^2 \leq \frac{\overline{W}^k}{b_k^{dk}}\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i|\sum_{l=0}^{k-1}k\binom{k-1}{l}X_i(\mathbb{X})^l\overline{\varphi}^{k-l-1}\lambda(\mathbb{X})^{k-l}|\hat{\theta} - \theta|\right)^2.$$

We can then conclude

$$\|\hat{g}_k - g_k\|^2 \leq \frac{\overline{W}^k k^2\lambda(\mathbb{X})^2}{b_k^{dk}}|\hat{\theta} - \theta|^2\left(\frac{1}{n}\sum_{i=1}^{n}|Y_i|\left(X_i(\mathbb{X}) + \overline{\varphi}\lambda(\mathbb{X})\right)^{k-1}\right)^2.$$

Lemma follows, since $\overline{\varphi}\lambda(\mathbb{X}) \geq 1$. $\square$

Denote for all $i,j \geq 0$

$$s_{i,j} = \mathbb{E}|Y|^i\left(X(\mathbb{X}) + \overline{\varphi}\lambda(\mathbb{X})\right)^j. \tag{7.3}$$

Moreover, $(V_n)_n$ is a sequence of real numbers, bigger than 1 and tending to infinity, to be tuned latter.

Lemma 7.2. *Let $l \geq 0$ be fixed. Then for all $x \in \mathbb{X}^k$,*

$$\mathbb{E}\left| Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right|^l \leq s_{l,lk} \frac{C^{lk}}{b_k^{ldk}}.$$

Moreover,

$$\mathbb{E}\left(Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \leq C^k \sqrt{s_{4,4k}} \left(k! \frac{V_n^2}{b_k^{dk}} + \frac{\mathrm{e}^{-uV_n/2}}{b_k^{2dk}} \right).$$

Proof. First observe that

$$\left| \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right| \leq \frac{\overline{W}^k}{b_k^{dk}} \sum_{l=0}^k \binom{k}{l} \overline{\varphi}^{k-l} \lambda(\mathbb{X})^{k-l} X(\mathbb{X})^l$$

$$\leq \frac{\overline{W}^k}{b_k^{dk}} \left(X(\mathbb{X}) + \overline{\varphi}\lambda(\mathbb{X}) \right)^k.$$

Hence,

$$\mathbb{E}\left| Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right|^l \leq \frac{\overline{W}^{lk} s_{l,lk}}{b_k^{ldk}}$$

Regarding the second inequality, we observe that

$$\mathbb{E}\left(Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \leq V_n^2 \mathbb{E}\left(\int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2$$

$$+ \mathbb{E}\left(Y \mathbf{1}_{|Y|>V_n} \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2.$$

By (3.1),

$$\mathbb{E}\left(\int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \leq k! \frac{C^k}{b_k^{dk}}.$$

Moreover, by the Cauchy−Schwarz Inequality, (2.6) and above,

$$\left[\mathbb{E}\left(Y \mathbf{1}_{|Y|>V_n} \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^2 \right]^2 \leq \mathbb{P}(|Y| > V_n)$$

$$+ \mathbb{E}\left(Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x-y) \tilde{X}^{\otimes k}(\mathrm{d}y) \right)^4$$

$$\leq C^k s_{4,4k} \frac{\mathrm{e}^{-uV_n}}{b_k^{4dk}}.$$

Puting all pieces together gives the result. $\square$

Lemma 7.3. *Let $k > \delta$ be fixed. Then,*

$$\mathbb{P}(\hat{\delta} = k) \leq C^k \frac{((2k)! + s_{1,k})^2}{n(\rho_k b_k^{dk/2})^2} + \frac{s_{2,2k}}{n((2k)!)^2} + \frac{C^k}{\rho_k^8} s_{8,8k} \max\left(\frac{n}{(nb_k^{dk})^8}, \frac{(k!)^4 V_n^8}{(nb_k^{dk})^4} + \frac{\mathrm{e}^{-2uV_n}}{(nb_k^{2dk})^4} \right).$$

Proof. Since $k > \delta$, we have with notation (7.1)

$$
\begin{aligned}
\mathbb{P}(\hat{\delta} = k) &\leq \mathbb{P}(\|\hat{g}_k\| > \rho_k) \\
&\leq \mathbb{P}(\|\hat{g}_k - g_k\| + \|g_k\| > \rho_k) \\
&\leq \mathbb{P}(2\|\hat{g}_k - g_k\| > \rho_k) + \mathbb{P}(2\|g_k\| > \rho_k).
\end{aligned}
\tag{7.4}
$$

Regarding the first term on the right-hand side, we observe that by Lemma 7.1 and (2.5):

$$
\begin{aligned}
\mathbb{P}(2\|\hat{g}_k - g_k\| > \rho_k) &\leq \mathbb{P}\left(C^k S_k \frac{|\hat{\theta} - \theta|}{b_k^{dk/2}} \geq \rho_k \right) \\
&\leq \mathbb{P}\left(C^k \big((2k)! + s_{1,k}\big)|\hat{\theta} - \theta| \geq \rho_k b_k^{dk/2} \right) \\
&\quad + \mathbb{P}\left(|S_k - s_{1,k}| \geq (2k)! \right) \\
&\leq C^k \frac{((2k)! + s_{1,k})^2}{n(\rho_k b_k^{dk/2})^2} + \frac{\mathbb{E}|S_k - s_{1,k}|^2}{((2k)!)^2}.
\end{aligned}
$$

Regarding the latter term on the right-hand side of inequality (7.4), since $k > \delta$, we have

$$
\mathbb{E}Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y)\tilde{X}^{\otimes k}(\mathrm{d}y) = 0,
$$

by (3.1). Thus, according to the Rosenthal Inequality (*e.g.* see Ibragimov and Sharakhmetov [7]), we get

$$
\begin{aligned}
\mathbb{E}\|g_k\|^8 &\leq \lambda(\mathbb{X})^{3k}\mathbb{E}\int_{\mathbb{X}^k} g_k^8 \mathrm{d}\lambda \\
&\leq C^k \int_{\mathbb{X}^k} \max\left(\frac{1}{n^7} D_{k,8}(x), \frac{1}{n^4}(D_{k,2})^4(x) \right) \mathrm{d}x,
\end{aligned}
$$

where for all $j \geq 0$ and $x \in \mathbb{X}^k$:

$$
D_{k,j}(x) = \mathbb{E}\left| Y \int_{\Delta_k} W_{b_k}^{\otimes k}(x - y)\tilde{X}^{\otimes k}(\mathrm{d}y) \right|^j.
$$

Finally, according to Lemma 7.2 and since $s_{4,4k}^2 \leq s_{8,8k}$, we obtain

$$
\mathbb{E}\|g_k\|^8 \leq C^k s_{8,8k} \max\left(\frac{n}{(nb_k^{dk})^8}, \frac{(k!)^4 V_n^8}{(nb_k^{dk})^4} + \frac{\mathrm{e}^{-2uV_n}}{(nb_k^{2dk})^4} \right).
$$

Lemma follows.   □

From now on, denote

$$
E_{k+1}^2 = 2(\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)\omega_{k+1}\lambda(\mathbb{X})^{k+2}\bar{f},
\tag{7.5}
$$

where $\omega_{k+1} = \int_{\mathbb{X}^{k+1}} |x| K^{\otimes k+1}(x)\mathrm{d}x$ and the other constants are defined in assumptions (2.8) and (2.9) on the model.

Lemma 7.4. *Let $k < \delta$ be fixed. Then,*

$$
\|\mathbb{E}g_{k+1}\| \geq m - E_{k+1}\sqrt{b_{k+1}}, \text{ and}
$$

$$
\mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 \leq C\left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{\mathrm{e}^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}} \right).
$$

Proof. By (3.1), for all $x \in \mathbb{X}^{k+1}$,

$$\mathbb{E}g_{k+1}(x) = \int_{\mathbb{X}^{k+1}} f_{k+1}(y)\overline{W_{b_{k+1}}^{\otimes k+1}(x-y)}\varphi_\theta^{\otimes k+1}(y)\mathrm{d}y$$

$$= \int_{\mathbb{X}^{k+1}} \left(f_{k+1}\varphi_\theta^{\otimes k+1}\right)(x-b_{k+1}z)\,\overline{W^{\otimes k+1}}(z)\mathrm{d}z.$$

By assumptions (2.1)−(2.8) on the model, we have

$$|(f_{k+1}\varphi_\theta^{\otimes k+1})(x-b_{k+1}z) - (f_{k+1}\varphi_\theta^{\otimes k+1})(x)| \leq (\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)b_{k+1}|z|.$$

Hence, since $\int_{\mathbb{X}} W \mathrm{d}\lambda = 1$,

$$\mathbb{E}g_{k+1} \geq f_{k+1}(x) - (\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)b_{k+1}\omega_{k+1}.$$

Then,

$$\|\mathbb{E}g_{k+1}\|^2 \geq \int_{\mathbb{X}^{k+1}} \left(f_{k+1}(x) - (\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)b_{k+1}\omega_{k+1}\right)^2 \mathrm{d}x$$

$$\geq \|f_{k+1}\|^2 - 2(\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)b_{k+1}\omega_{k+1}\int_{\mathbb{X}^{k+1}} f_{k+1}\mathrm{d}\lambda$$

$$\geq \|f_{k+1}\|^2 - 2(\overline{\varphi}^{k+1}\gamma_2 + k\bar{f}\overline{\varphi}^k)b_{k+1}\omega_{k+1}\lambda(\mathbb{X})^{k+1}\|f_{k+1}\|.$$

First part of the lemma follows, since $\|f_{k+1}\| \geq m$ and $|f_{k+1}| \leq \bar{f}$. Moreover, the second part is straightforward from Lemma 7.2, since

$$\mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 = \int_{\mathbb{X}^{k+1}} \mathrm{var}\big(g_{k+1}(x)\big)\mathrm{d}x$$

$$= \frac{1}{n}\int_{\mathbb{X}^{k+1}} \mathrm{var}\left(Y\int_{\Delta_{k+1}} W_{b_{k+1}}^{\otimes k+1}(x-y)\tilde{X}^{\otimes k+1}(\mathrm{d}y)\right)\mathrm{d}x$$

$$\leq \frac{1}{n}\int_{\mathbb{X}^{k+1}} \mathbb{E}\left(Y\int_{\Delta_{k+1}} W_{b_{k+1}}^{\otimes k+1}(x-y)\tilde{X}^{\otimes k+1}(\mathrm{d}y)\right)^2 \mathrm{d}x. \qquad \square$$

Lemma 7.5. *We have, for all $k < \delta$:*

$$\mathbb{P}(\hat{\delta} = k) \leq C\left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{e^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}}\right).$$

Proof. Since $\|f_{k+1}\| \neq 0$,

$$\mathbb{P}(\hat{\delta} = k) \leq \mathbb{P}(\|\hat{g}_{k+1}\| \leq \rho_{k+1})$$

$$\leq \mathbb{P}(\|\hat{g}_{k+1} - g_{k+1}\| + \|g_{k+1} - \mathbb{E}g_{k+1}\| \geq \|\mathbb{E}g_{k+1}\| - \rho_{k+1})$$

$$\leq \mathbb{P}\left(\|\hat{g}_{k+1} - g_{k+1}\| \geq \frac{\|\mathbb{E}g_{k+1}\| - \rho_{k+1}}{2}\right)$$

$$+ \mathbb{P}\left(\|g_{k+1} - \mathbb{E}g_{k+1}\| \geq \frac{\|\mathbb{E}g_{k+1}\| - \rho_{k+1}}{2}\right). \tag{7.6}$$

According to Lemma 7.1 and using the lower bound obtained in Lemma 7.4, we find that the first term on the right-hand side of (7.6), denoted by p_1, is upper bounded by

$$\mathbb{P}\left(C^k S_{k+1}\frac{|\hat{\theta} - \theta|}{b_{k+1}^{d(k+1)/2}} \geq \frac{1}{2}\left(m - E_{k+1}\sqrt{b_{k+1}} - \rho_{k+1}\right)\right).$$

Since by assumption b_{k+1}, ρ_{k+1} tend to 0 as n tends to $+\infty$, for all $k < \delta$:

$$p_1 \leq \mathbb{P}\left(S_{k+1}|\hat{\theta} - \theta| \geq b_{k+1}^{d(k+1)/2}I\right),$$

for some $I > 0$ that does not depend on n and $k < \delta$. Thus, by the Markov Inequality,

$$\begin{aligned}
p_1 &\leq \mathbb{P}\left((1 + s_{1,k+1})|\hat{\theta} - \theta| \geq b_{k+1}^{d(k+1)/2}I\right) + \mathbb{P}\left(|S_{k+1} - s_{1,k+1}| \geq 1\right) \\
&\leq \frac{(1 + s_{1,k+1})^2}{I^2} \frac{\mathbb{E}|\hat{\theta} - \theta|^2}{b_{k+1}^{d(k+1)}} + \mathbb{E}(S_{k+1} - s_{1,k+1})^2 \\
&\leq \frac{C}{nb_{k+1}^{d(k+1)}} + \frac{C}{n},
\end{aligned} \qquad (7.7)$$

by assumption (2.5) on the model, and since $s_{2,2(k+1)} \leq C$. In a similar fashion, regarding the latter term on the right-hand side of (7.6), further denoted by p_2, we obtain with Lemma 7.4:

$$p_2 \leq \frac{4}{I^2}\mathbb{E}\|g_{k+1} - \mathbb{E}g_{k+1}\|^2 \leq C\left(\frac{V_n^2}{nb_{k+1}^{d(k+1)}} + \frac{\mathrm{e}^{-uV_n/2}}{nb_{k+1}^{2d(k+1)}}\right).$$

We conclude the proof combining (7.6), (7.7) and above, since $V_n \geq 1$. $\qquad \square$

7.2. Proof of Theorem 4.1

First observe that for all $k \geq 1$, $\hat{\delta}$ is independent from $\hat{r}_k$, so that

$$\begin{aligned}
\mathbb{E}\left(\hat{r}_{\hat{\delta}}(X) - r(X)\right)^2 &\leq \mathbb{E}\left(\hat{r}_\delta(X) - r(X)\right)^2 + \sum_{k<\delta} \mathbb{E}\left(\hat{r}_k(X) - r(X)\right)^2 \mathbb{P}(\hat{\delta} = k) \\
&\quad + \sum_{k>\delta} \mathbb{E}\left(\hat{r}_k(X) - r(X)\right)^2 \mathbb{P}(\hat{\delta} = k).
\end{aligned} \qquad (7.8)$$

By Lemma 5.3, it is clear that for $k < \delta$,

$$\mathbb{E}\left(\hat{r}_k(X) - r(X)\right)^2 \leq C.$$

Moreover, following the arguments of the proof of Theorem 3.1 (see Sect. 5.2), we can prove that for all $k \geq \delta$ (recall that $C > 0$ does not depend on k nor n):

$$\mathbb{E}\left(\hat{r}_k(X) - r(X)\right)^2 \leq C^k \left(\frac{T_n^2}{n}\right)^{2/(2+dk)}.$$

Thus, by Theorem 3.1, (7.8) and above,

$$\begin{aligned}
\mathbb{E}\left(\hat{r}_{\hat{\delta}}(X) - r(X)\right)^2 &\leq C\left(\frac{T_n^2}{n}\right)^{2/(2+d\delta)} + C\sum_{k<\delta} \mathbb{P}(\hat{\delta} = k) \\
&\quad + \sum_{k>\delta} C^k \mathbb{P}(\hat{\delta} = k).
\end{aligned} \qquad (7.9)$$

Let ρ_k and b_k be defined in Theorem 4.1, and let $V_n = 2(\ln n)^\xi/u$, where $\xi > 1$. Then by Lemma 7.5, we get for $k < \delta$:

$$\mathbb{P}(\hat{\delta} = k) \leq C\frac{(\ln n)^{2\xi}}{n^{1-\beta/2}}.$$

Provided $\varepsilon > d\delta$, $\xi \leq (2+\varepsilon)/(2+d\delta)$ and $\beta \leq 2d\delta/(2+d\delta)$, we have

$$\sum_{k<\delta} \mathbb{P}(\hat{\delta} = k) \leq C\left(\frac{T_n^2}{n}\right)^{2/(2+d\delta)}. \tag{7.10}$$

Note that when $\beta < 1$, the previous condition on β holds if $d\delta \geq 2$. Our task now is to bound $\mathbb{P}(\hat{\delta} = k)$ when $k > \delta$. First, we observe that for all $i, j \geq 0$:

$$s_{i,j} \leq \left(\mathbb{E}Y^{2i}\right)^{1/2}\left[\left(\mathbb{E}\left(X(\mathbb{X})^{2j}\right)\right)^{1/2} + \overline{\varphi}^j\lambda(\mathbb{X})^j\right]$$
$$\leq \left(\mathbb{E}Y^{2i}\right)^{1/2}\left[\mathrm{e}^{\overline{\varphi}\lambda(\mathbb{X})}\sqrt{(2j)!} + \overline{\varphi}^j\lambda(\mathbb{X})^j\right]$$

according to the Cauchy−Schwarz Inequality. Consequently, by Lemma 7.3, if $k > \delta$:

$$\mathbb{P}(\hat{\delta} = k) \leq \frac{C^k}{((2k)!)^2 n^{\alpha+\beta/2}} + \frac{C^k\sqrt{(4k)!}}{n((2k)!)^2}$$
$$+ \frac{C^k\sqrt{(16k)!}}{((2k)!)^{16}}\max\left(\frac{1}{n^{4\alpha+3}}, \frac{(k!)^4 V_n^8}{n^{4\alpha+2\beta}} + \frac{\mathrm{e}^{-2uV_n}}{n^{4\alpha}}\right).$$

Noticing that by the Stirling Formula,

$$\sum_{k\geq 1} C^k\left(\frac{\sqrt{(4k)!}}{((2k)!)^2} + \frac{(k!)^4\sqrt{(16k)!}}{((2k)!)^{16}}\right) < \infty,$$

we deduce, whenever $2\alpha + \beta > 1/(2+d\delta)$:

$$\sum_{k>\delta} \mathbb{P}(\hat{\delta} = k) \leq C\left(\frac{T_n^2}{n}\right)^{2/(2+d\delta)}.$$

Theorem is now a straightforward consequence of (7.9), (7.10) and above. $\qquad\square$

References

[1] G. Biau, F. Cérou and A. Guyader, Rates of convergence of the functional k-nearest neighbor estimate. *IEEE Trans. Inform. Theory* **56** (2010) 2034–2040.

[2] L. Birgé, Approximation dans les espaces métriques et théorie de l'estimation. *L. Z. Wahrscheinlichkeitstheorie verw Gebiete* **65** (1983) 181–237.

[3] B. Cadre and L. Truquet, Nonparametric regression estimation onto a Poisson point process covariate. *ESAIM: PS* **19** (2015) 251–267.

[4] G. Chagny and A. Roche, Adaptive estimation in the functional nonparametric regression model. *J. Multivariate Anal.* **146** (2016) 105–118.

[5] L. Györfi, M. Kohler, A. Krzyzak and H. Walk, A distribution-free theory of nonparametric regression. Springer Science and Business Media (2002).

[6] L. Horváth and P. Kokoszka, *Inference for functional data with applications*. Vol. 200. Springer Science and Business Media (2012).

[7] R. Ibragimov and Sh. Sharakhmetov, The exact constant in the Rosenthal Inequality for random variables with mean zero. *Theory Probab. Appl.* **46** (1999) 127–131.

[8] K. Itô, Spectral type of the shift transformation of differential processes with stationary increments. *Trans. Amer. Math. Soc.* (1956) 253–263.

[9] M. Kohler, A. Krzyżak and H. Walk, Optimal global rates of convergence for nonparametric regression with unbounded data. *J. Statist. Plann. Inference* **139** (2009) 1286–1296.

[10] G. Last and M.D. Penrose, Poisson process Fock space representation, chaos expansion and covariance inequalities. *Probab. Theory Relat. Fields* **150** (2011) 663–690.

[11] A. Mas, Lower bound in regression for functional data by representation of small ball probabilities. *Electr. J. Statist.* **6** (2012) 1745–1778.

[12] D. Nualart and J. Vives, Anticipative calculus for the Poisson process based on the Fock space, Séminaire de Probabilitées XXIV. In: *Lect. Notes Math.* (1990) 154–165.

[13] J.O. Ramsay and B.W. Silverman, Functional data analysis. Springer Science & Business Media (2005).

[14] A.B. Tsybakov, Introduction to nonparametric estimation. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. *Springer Series in Statistics.* Springer, New York (2009).