

Nonparametric regression estimation onto a Poisson point process covariate

Benoît CADRE*
IRMAR, ENS Rennes, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
cadre@ens-rennes.fr

Lionel TRUQUET
IRMAR, Ensai, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman, 35170 Bruz, France
truquet@ensai.fr

Abstract

Let Y be a real random variable and X be a Poisson point process. We investigate rates of convergence of a nonparametric estimate $\hat{r}(x)$ of the regression function $r(x) = \mathbb{E}(Y|X = x)$, based on n independent copies of the pair (X, Y) . The estimator \hat{r} is constructed using a Wiener-Itô decomposition of $r(X)$. In this infinite-dimensional setting, we first obtain a finite sample bound on the expected squared difference $\mathbb{E}(\hat{r}(X) - r(X))^2$. Then, under a condition ensuring that the model is genuinely infinite-dimensional, we obtain the exact rate of convergence of $\ln \mathbb{E}(\hat{r}(X) - r(X))^2$.

Index Terms — Regression estimation, Poisson point process, Wiener-Itô decomposition, Rates of convergence.

AMS 2010 Classification — 62G05, 62G08

*corresponding author

1 Introduction

1.1 Functional regression estimation

Let \mathcal{S} be a measurable space, and let the data $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent $\mathcal{S} \times \mathbb{R}$ -valued random variables with the same distribution as a generic pair (X, Y) such that $\mathbb{E}|Y| < \infty$. In the regression estimation problem, the goal is to estimate the regression function $r(X) = \mathbb{E}(Y|X)$ using the data.

In the classical setting, each covariate X_i is supposed to be a collection of numerical experiments represented by a finite-dimensional vector. Thus, to date, most of the results pertaining to regression estimation have been reported in the finite-dimensional case where $\mathcal{S} = \mathbb{R}^d$. We refer the reader to the book by Györfi et al [5] for a comprehensive introduction to the subject and an overview of most standard methods in \mathbb{R}^d .

However, in an increasing number of practical applications, input data items take more complicated forms. In the functional data analysis, \mathcal{S} is a set of curves and, in this context, regression estimation has many applications in a wide class of problems, among with speech recording, analysis of patients visits in a hospital, price of an option... Last few years have witnessed important developments in both the theory and practise of functional data analysis, and many traditional statistical tools have been adapted to handle functional inputs. The book by Ramsay and Silverman [11] provides a presentation of this area.

In the infinite-dimensional setting, the regression problem is faced with new challenges which requires changing methodology. Curiously, despite a huge research activity in the area of infinite-dimensional data analysis, few attempts have been made to connect it with the rich theory of stochastic processes that both provides a wide class of models and powerful tools. This approach, based on a use of stochastic process theory for the benefit of nonparametric estimation, has been studied in the fairly closed problem of supervised classification; in this direction, we refer the reader to the recent papers by Baillo et al [3], Biau et al [2], Cadre [4].

Among the classical models from the theory of time-dependent stochastic processes, the case of Poisson processes is of great interest. Here, \mathcal{S} is the set of counting paths on a subset of \mathbb{R}_+ . In epidemiology for example, the observed curve X_i represent the dates of the patient visits to the doctor and the variable response Y_i provide quantitative information on the health status of the patient.

More generally, we consider in this paper the case of a Poisson point process covariate, which corresponds to the situation where each measurement report the

locations of every individual event. In this setting, the state space \mathcal{S} is identified to the so-called Poisson space over some measurable space \mathbb{X} , i.e.

$$\mathcal{S} = \left\{ \sum_{i=1}^n \delta_{x_i}, n \in \mathbb{N} \text{ and } x_i \in \mathbb{X} \right\},$$

where δ_x stands for the Dirac measure on x .

1.2 Poisson point process regression estimation

In the sequel, the covariate X is a Poisson point process (see [7]) on a domain $\mathbb{X} \subset \mathbb{R}^d$ with mean measure μ , where μ is a σ -finite measure on the Borel σ -field \mathcal{X} of \mathbb{X} . As seen before, this means X belongs to the space of integer-valued σ -finite measures on \mathbb{X} and satisfies:

- for any $A \in \mathcal{X}$, the number X_A of points of X lying in A has a Poisson distribution with parameter $\mu(A)$;
- for any family of disjoint sets $A_1, \dots, A_\ell \in \mathcal{X}$, $X_{A_1}, \dots, X_{A_\ell}$ are independent random variables.

In the case $\mathbb{X} = \mathbb{R}$, Itô's famous chaos expansion (see [6], [13]) says that every square integrable and $\sigma(X)$ -measurable random variable can be decomposed as a sum of multiple stochastic integrals, called chaos. This result has been generalized by Nualart and Vives [10], and more recently by Last and Penrose [8].

Now recall some basic facts about chaotic decomposition in the Poisson space. Fix $k \geq 1$. Provided $g \in \mathbb{L}^2(\mu^{\otimes k})$, we can define the k -th chaos $I_k(g)$ associated with g , namely

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

where $\Delta_k = \{x \in \mathbb{X}^k : x_i \neq x_j \text{ for all } i \neq j\}$. Interestingly, if $g \in \mathbb{L}^2(\mu^{\otimes k})$ and $h \in \mathbb{L}^2(\mu^{\otimes \ell})$ for $k, \ell \geq 1$, we have

$$\mathbb{E}I_k(g)I_\ell(h) = k! \int_{\mathbb{X}^k} \bar{g} \bar{h} d\mu^{\otimes k} \mathbf{1}_{\{k=\ell\}} \text{ and } \mathbb{E}I_k(g) = 0, \quad (1.2)$$

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

$$\bar{g}(x_1, \dots, x_k) = \frac{1}{k!} \sum_{\sigma} g(x_{\sigma(1)}, \dots, x_{\sigma(k)}),$$

the sum being taken over all permutations $\sigma = (\sigma(1), \dots, \sigma(k))$ of $\{1, \dots, k\}$, and similarly for \bar{h} . In particular, note that $I_k(g)$ is a square integrable random variable. In Nualart and Vives ([10], p. 160), 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

$$r(X) = \mathbb{E}Y + \sum_{k \geq 1} \frac{1}{k!} I_k(f_k), \quad (1.3)$$

where equality holds in \mathbb{L}^2 , provided $\mathbb{E}Y^2 < \infty$. In the above formula, each f_k is an element of $\mathbb{L}_{\text{sym}}^2(\mu^{\otimes k})$ –the subset of symmetric functions in $\mathbb{L}^2(\mu^{\otimes k})$ –, and the decomposition is defined in an unique way. Last and Penrose [8] proved that each f_k can be expressed as a difference operator of order k .

Based on independent copies of (X, Y) , we shall construct a nonparametric estimate \hat{r} of r with the help of decomposition (1.3). Next section is devoted to the model, and to the construction and statistical properties of \hat{r} . In particular, we obtain a finite sample bound on the mean squared error $\mathbb{E}(\hat{r}(X) - r(X))^2$. Moreover, we prove that if the model is genuinely infinite-dimensional in the sense that $\inf_{k \geq 1} \|f_k\|_{\mathbb{L}^2(\mu^{\otimes k})} > 0$ then, under some regularity conditions:

$$\lim_{n \rightarrow \infty} \frac{\ln \mathbb{E}(\hat{r}(X) - r(X))^2}{\sqrt{\ln n \ln_2 n}} = -\sqrt{\frac{\alpha}{2d}},$$

for some $\alpha \in]0, 1[$. Last two sections contains proofs.

2 Regression estimate

2.1 Model and heuristic of the estimate

Basic assumptions on the model. We assume throughout that \mathbb{X} is a compact set, say $\mathbb{X} \subset [-M, M]^d$, and the mean measure μ has a density φ with respect to the Lebesgue measure λ on \mathbb{X} that is, $\varphi : \mathbb{X} \rightarrow \mathbb{R}_+$ is such that for all Borel set $A \subset \mathbb{X}$:

$$\mathbb{E}X_A = \int_A \varphi d\lambda.$$

Heuristics. In view of a short presentation of the heuristic of the estimate of r , we assume for simplicity that φ is a known positive function. Suppose also that

the intensity φ and the f_k 's defined by (1.3) are bounded functions. Let W be a bounded density on \mathbb{X} , $h = h(n) > 0$ and for all $x \in \mathbb{X}$:

$$W_h(x) = \frac{1}{h^d} W\left(\frac{x}{h}\right). \quad (2.1)$$

For a real-valued function g defined on \mathbb{X} , the notation $g^{\otimes k}$ denotes the real-valued function on \mathbb{X}^k such that

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

By relations (1.2) and (1.3), we have for all $x \in \mathbb{X}^k$ and $k \geq 1$:

$$\begin{aligned} \mathbb{E}YI_k(W_h^{\otimes k}(x - \cdot)) &= \mathbb{E}r(X)I_k(W_h^{\otimes k}(x - \cdot)) \\ &= \sum_{\ell \geq 1} \frac{1}{\ell!} \mathbb{E}I_\ell(f_\ell)I_k(W_h^{\otimes k}(x - \cdot)) \\ &= \int_{\mathbb{X}^k} f_k \bar{W}_h^{\otimes k}(x - \cdot) \varphi^{\otimes k} d\lambda^{\otimes k}, \end{aligned}$$

where $\bar{W}_h^{\otimes k}(x - \cdot)$ is the symmetrization of the function $W_h^{\otimes k}(x - \cdot)$. Since f_k is a symmetric function, we can write

$$\mathbb{E}YI_k(W_h^{\otimes k}(x - \cdot)) = \int_{\mathbb{X}^k} f_k W_h^{\otimes k}(x - \cdot) \varphi^{\otimes k} d\lambda^{\otimes k}.$$

Thus, under smoothness assumptions on φ and f_k , the right-hand side converges to $f_k(x) \varphi^{\otimes k}(x)$, provided $h \rightarrow 0$. From the left-hand side, we thus deduce that a kernel-type estimator of $f_k(x)$ based on independent copies $(X_1, Y_1), \dots, (X_n, Y_n)$ of (X, Y) is

$$\tilde{f}_k(x) = \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k} \frac{W_h^{\otimes k}(x - \cdot)}{\varphi^{\otimes k}(x)} d(X_i - \mu)^{\otimes k}.$$

By (1.1), the k -th chaos $I_k(f_k)$ is thus estimated by

$$\tilde{I}_k = \int_{\Delta_k} \tilde{f}_k(x) d(X - \mu)^{\otimes k}, \quad (2.2)$$

which, by (1.3), gives the estimator \tilde{r} of r defined by

$$\tilde{r}(X) = \bar{Y}_n + \sum_{k=1}^N \frac{1}{k!} \tilde{I}_k, \quad (2.3)$$

where $N = N(n)$ tends to infinity and, as usual,

$$\bar{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i.$$

However, this construction requires the unrealistic assumption that φ is known. Thus we shall first present a nonparametric estimator of φ , then we shall adapt the previous idea to this context.

2.2 Construction of the estimator

Estimation of φ . Construction of the nonparametric estimate of φ is based on Mecke's Formula (see [9]), which states in particular that

$$\mathbb{E} \int_{\mathbb{X}} \gamma dX = \int_{\mathbb{X}} \gamma d\mu = \int_{\mathbb{X}} \gamma \varphi d\lambda,$$

provided $\gamma : \mathbb{X} \rightarrow \mathbb{R}$ is in $\mathbb{L}^1(\mu)$. With this respect, we define the nonparametric estimator $\hat{\varphi}$ of φ by

$$\hat{\varphi}(x) = \frac{1}{n} \sum_{j=1}^n \int_{\mathbb{X}} K_b(x, y) X_j(dy), \quad x \in \mathbb{X},$$

where, for all $x = (x_1, \dots, x_d)$ and $u = (u_1, \dots, u_d) \in \mathbb{X}$,

$$K_b(x, u) = \prod_{j=1}^d (J_b(x_j - u_j) + J_b(2M + x_j + u_j) + J_b(2M - x_j - u_j)). \quad (2.4)$$

In the above formula, J is a continuous and symmetric density on $[-1, 1]$ and, for the bandwidth $b = b(n) > 0$ such that $b \rightarrow 0$ as $n \rightarrow \infty$:

$$J_b(y) = \frac{1}{b} J\left(\frac{y}{b}\right), \quad y \in [-b, b].$$

This particular construction is classical in order to avoid bias on the boundary of \mathbb{X} (see p. 30 in the book by Silverman [12]). Then, we define for all $i = 1, \dots, n$ the leave-one-out nonparametric estimator $\hat{\varphi}_i$ of φ by

$$\begin{aligned} \hat{\varphi}_i(x) &= \hat{\varphi}(x) - \frac{1}{n} \int_{\mathbb{X}} K_b(x, y) X_i(dy) \\ &= \frac{1}{n} \sum_{j=1, j \neq i}^n \int_{\mathbb{X}} K_b(x, y) X_j(dy), \end{aligned}$$

for $x \in \mathbb{X}$. Leave-one-out procedure is only considered here for technical matters.

Chaos estimate. Now consider the vanishing sequences $\rho = \rho(n) > 0$ and $h = h(n) > 0$. Following (2.2), the leave-one-out estimator of the k -th chaos $I_k(f_k)$ is \hat{I}_k , such that

$$\hat{I}_k = \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k} \left[\int_{\Delta_k} \frac{W_h^{\otimes k}(x - \cdot)}{(\hat{\phi}_i + \rho)^{\otimes k}(x)} (dX_i - \hat{\phi}_i d\lambda)^{\otimes k} \right] (dX - \hat{\phi}_i d\lambda)^{\otimes k}(x),$$

where W_h is defined by (2.1). In the sequel, we assume for simplicity that W has a compact support.

Regression estimate. Finally, following the idea drawn by (2.3), the estimator \hat{r} of r is

$$\hat{r}(X) = \bar{Y}_n + \sum_{k=1}^N \frac{1}{k!} \hat{I}_k,$$

where $N = N(n)$ tends to infinity.

2.3 Result

In the sequel, $\|\cdot\|$ is the euclidean norm on any \mathbb{R}^p . We assume that (X, Y) is independent from the sample $(X_1, Y_1), \dots, (X_n, Y_n)$. Now introduce the assumptions on the model.

H1. Y is a bounded random variable.

H2. $\inf_{\mathbb{X}} \varphi > 0$ and there exists $L_1 > 0$ such that for all $x, y \in \mathbb{X}$:

$$|\varphi(x) - \varphi(y)| \leq L_1 \|x - y\|.$$

H3. There exists a constant $L_2 > 0$ such that for all $k \geq 1$ and $x, y \in \mathbb{X}^k$:

$$|f_k(x) - f_k(y)| \leq L_2^k \|x - y\|.$$

H4. There exist two constants $C_1, C_2 > 0$ such that nb^{d+2} , nh^{dN+1} and $nb^d \rho^3 / N$ are bounded below by C_1 , and $b^2 + \rho \leq C_2 h^{dN+1}$.

Assumptions **H1-H3** are classical in nonparametric estimation. Moreover, it is easily seen that **H4** holds for $b = n^{-\gamma}$, $\rho = n^{-\beta}$ and h, N defined by formulas (2.6)

below, under the additional constraints that $0 < \gamma < 1/(d+2)$, $0 < 3\beta < 1 - d\gamma$ and $\alpha < \min(1, \beta, 2\gamma)$. Note that last constraint is sufficient for $b^2 + \rho \leq C_2 h^{dN+1}$ to hold because, under the condition $u_N/(N \ln N) \rightarrow 0$ as $n \rightarrow \infty$:

$$\ln h^{dN+1} = -\alpha \ln n + o(\ln n).$$

Moreover, since

$$\ln nh^{dN+1} = (1 - \alpha) \ln n + o(\ln n),$$

we see that $\alpha < 1$ is a sufficient condition for $nh^{dN+1} \geq C_1$ to hold.

For notational simplicity, we set $\ln_2 n = \ln(\ln n)$.

Theorem 2.1. *Assume that **H1-H4** hold. Then, there exists $C \geq 1$ such that*

$$\mathbb{E}(\hat{r}(X) - r(X))^2 \leq C^N \left(h + \frac{1}{N!} \right). \quad (2.5)$$

In particular, for the following choices:

$$N = \left\lceil \sqrt{\frac{2\alpha \ln n}{d \ln_2 n}} \right\rceil, \quad h = e^{-N \ln N - u_N}, \quad (2.6)$$

where $\alpha > 0$ and $u_N \geq 0$ is such that $u_N/(N \ln N) \rightarrow 0$ as $n \rightarrow \infty$, we have:

$$\limsup_{n \rightarrow \infty} \frac{\ln \mathbb{E}(\hat{r}(X) - r(X))^2}{\sqrt{\ln n \ln_2 n}} \leq -\sqrt{\frac{\alpha}{2d}}.$$

Roughly, Theorem 2.1 entails that the rate of convergence of the mean squared error between $\hat{r}(X)$ and $r(X)$ is at least

$$\exp\left(-\sqrt{\frac{\alpha}{2d} \ln n \ln_2 n}\right).$$

Note that this rate is obtained for that bandwidths h that satisfy, as $n \rightarrow \infty$:

$$\ln h \sim -\sqrt{\frac{\alpha}{2d} \ln n \ln_2 n}.$$

Hence, the general finding here is that the rate of convergence of the mean squared error is much slower than the traditional finite-dimensional rate (see [5]), but faster than the rates obtained by Biau et al [1] in the infinite-dimensional setting. This,

of course, is explained by the fact that we fully exploit the particular nature of the covariate via the chaotic decomposition of square integrable $\sigma(X)$ -measurable random variables whereas Biau et al [1] utilize a k -nearest neighbor estimator whose construction does not depend on the law of X .

Our task is now to prove that the rate of Theorem 2.1 is optimal in a genuine infinite-dimensional setting. With this goal, we see that it is necessary to strengthen the conditions on the model. Indeed, if all the f_k 's for $k \geq k_0$ have a null $\mathbb{L}_{\text{sym}}^2(\mu^{\otimes k})$ -norm, then r can be decomposed into a finite sum of chaos (see 1.3), and hence we are faced with a finite-dimensional estimation problem, for which the rate of theorem 2.1 may not be optimal. To avoid this situation, we assume that all the f_k 's have a $\mathbb{L}^2(\mu^{\otimes k})$ -norm greater than a positive constant.

Theorem 2.2. *Assume that **H1-H4** hold, and $\inf_{k \geq 1} \|f_k\|_{\mathbb{L}^2(\mu^{\otimes k})} > 0$. If N and h are given by (2.6) with the additional assumption on u_N that $u_N/N \rightarrow \infty$, we have*

$$\lim_{n \rightarrow \infty} \frac{\ln \mathbb{E}(\hat{r}(X) - r(X))^2}{\sqrt{\ln n \ln_2 n}} = -\sqrt{\frac{\alpha}{2d}}.$$

To the best of our knowledge, this is the first exact rate in the regression estimation problem with a Poisson point process covariate. However, it is an open problem to know whether this rate is optimal over the whole class of regression estimates.

3 Proof of Theorems 2.1 and 2.2

In the sequel, $\kappa > 0$ is such that $\sup_{\mathbb{X}} \varphi \sup_{\mathbb{R}^{2d}} K \leq \kappa$ and we assume for simplicity that constants C_1, C_2 of assumption **H4** are equal to 1. Moreover, we assume that ρ, b and h are smaller than 1 (recall that they vanish as n tends to infinity). Finally, we let for all $k \geq 1$:

$$\beta_k = \rho^{-k} \exp\left(-\frac{nb^d \rho^2}{4\kappa}\right).$$

We start the section with the following result, whose proof is presented later in this section.

Lemma 3.1. *Assume that assumptions **H1-H3** hold and $nb^{d+2} \geq 1$. Then, there exists $C \geq 1$ such that for all $k \geq 1$:*

$$\mathbb{E}(\hat{I}_k - I_k(f_k))^2 \leq C^k \left((k!)^2 \frac{b^2 + \beta_{2k}}{h^{dk}} + h + \frac{k!}{nh^{dk}} \right).$$

Proof of Theorem 2.1. In the sequel, $C \geq 1$ denotes a constant whose value may change from line to line. According to Jensen Inequality and Lemma 3.1:

$$\begin{aligned} \mathbb{E}\left(\sum_{k=1}^N \frac{1}{k!} (\hat{I}_k - I_k(f_k))\right)^2 &\leq C^N \sum_{k=1}^N \left(\frac{b^2 + \beta_{2k}}{h^{dk}} + \frac{h}{(k!)^2} + \frac{1}{k!nh^{dk}}\right) \\ &\leq C^N \left(\frac{b^2 + \beta_{2N}}{h^{dN}} + h + \frac{1}{nh^{dN}}\right). \end{aligned} \quad (3.1)$$

Moreover, according to Theorem 4.2 in Last and Penrose [8],

$$\mathbb{E}\left(\sum_{k \geq N+1} \frac{I_k(f_k)}{k!}\right)^2 \leq \frac{1}{N!} \mathbb{E} \int_{\mathbb{X}^{N+2}} [D_x^{N+2} r(X)]^2 \varphi^{\otimes(N+2)}(x) \lambda^{\otimes(N+2)}(dx), \quad (3.2)$$

where for all $x \in \mathbb{X}^{N+2}$, D_x^{N+2} denotes the difference operator of order $N+2$, that is, if δ_z is the Dirac mass on z :

$$D_x^{N+2} r(X) = \sum_{S \subset \{1, \dots, N+2\}} (-1)^{N+2-|S|} r(X + \sum_{s \in S} \delta_{x_s}).$$

In the formula above, $|S|$ is the number of elements of S . But $|r|$ is bounded since Y is bounded under **H1**. Hence, $|D_x^{N+2} r(X)| \leq C2^{N+2}$ and, by (3.2):

$$\mathbb{E}\left(\sum_{k \geq N+1} \frac{I_k(f_k)}{k!}\right)^2 \leq \frac{C^N}{N!}.$$

Putting all pieces together, we deduce from (3.1) and above that

$$\mathbb{E}(\hat{r}(X) - r(X))^2 \leq C^N \left(\frac{1}{n} + \frac{b^2 + \beta_{2N}}{h^{dN}} + h + \frac{1}{nh^{dN}} + \frac{1}{N!}\right),$$

because $\mathbb{E}(\mathbb{E}Y - \bar{Y}_n)^2 \leq 1/n$. Now, under condition **H4** which in particular states that $nb^d \rho^3 / N \geq 1$, we have for n large enough:

$$\beta_{2N} \leq \rho^{-2N} e^{3N \ln \rho} \leq \rho. \quad (3.3)$$

Moreover, since $nh^{dN+1} \geq 1$ and $b^2 + \rho \leq h^{dN+1}$:

$$\mathbb{E}(\hat{r}(X) - r(X))^2 \leq C^N \left(h + \frac{1}{N!}\right), \quad (3.4)$$

hence the first part of the theorem.

Now consider the case where N and h are given by (2.6). We have

$$C^N h = e^{N \ln C - N \ln N - u_N} \leq e^{-N \ln N + N \ln C}. \quad (3.5)$$

Furthermore, according to the Stirling Formula,

$$\frac{C^N}{N!} \leq e^{N \ln C + N - N \ln N}.$$

Hence by (3.4), we have for all $\varepsilon > 0$:

$$\begin{aligned} \ln \mathbb{E}(\hat{r}(X) - r(X))^2 &\leq \ln(e^{-N \ln N + N \ln C} + e^{-N \ln N + N(1 + \ln C)}) \\ &\leq -N \ln N + \ln(e^{N \ln C} + e^{N(1 + \ln C)}). \end{aligned}$$

Moreover, by the very definition of N given in (2.6):

$$N \ln N = \left[\sqrt{\frac{2\alpha \ln n}{d \ln_2 n}} \right] \ln \left[\sqrt{\frac{2\alpha \ln n}{d \ln_2 n}} \right] = \sqrt{\frac{\alpha}{2d}} \ln n \ln_2 n (1 + o(1)). \quad (3.6)$$

Putting all pieces together yield

$$\limsup_{n \rightarrow \infty} \frac{\ln \mathbb{E}(\hat{r}(X) - r(X))^2}{\sqrt{\ln n \ln_2 n}} \leq -\sqrt{\frac{\alpha}{2d}},$$

which is the desired result. \square

Proof of Theorem 2.2. For simplicity, we assume that $\inf_{k \geq 1} \|f_k\|_{\mathbb{L}^2(\mu^{\otimes k})} \geq 1$. Then, by the triangle inequality and (1.2):

$$\begin{aligned} \sqrt{\mathbb{E}(\hat{r}(X) - r(X))^2} &= \sqrt{\mathbb{E} \left[\sum_{k=1}^N \frac{1}{k!} (\hat{I}_k - I_k(f_k)) - \sum_{k \geq N+1} \frac{1}{k!} I_k(f_k) \right]^2} \\ &\geq \sqrt{\mathbb{E} \left(\sum_{k \geq N+1} \frac{1}{k!} I_k(f_k) \right)^2} - \sqrt{\mathbb{E} \left(\sum_{k=1}^N \frac{1}{k!} (\hat{I}_k - I_k(f_k)) \right)^2} \\ &\geq \frac{1}{\sqrt{(N+1)!}} - \sqrt{\mathbb{E} \left(\sum_{k=1}^N \frac{1}{k!} (\hat{I}_k - I_k(f_k)) \right)^2}. \end{aligned}$$

Applying successively (3.1), (3.3) and **H4**, we first deduce that

$$\mathbb{E} \left(\sum_{k=1}^N \frac{1}{k!} (\hat{I}_k - I_k(f_k)) \right)^2 \leq C^N h = e^{-N \ln N + N \ln C - u_N}.$$

Moreover, by the Stirling Formula:

$$\frac{1}{(N+1)!} \geq e^{-(N+3/2) \ln(N+1) + N}.$$

Consequently, $\mathbb{E}(\hat{r}(X) - r(X))^2$ is greater than

$$\left(e^{-(N/2+3/4) \ln(N+1) + N/2} - e^{-(N/2) \ln N + (N/2) \ln C - u_N/2} \right)^2.$$

Then, using the condition $u_N/N \rightarrow \infty$, we get with easy calculations that

$$\ln \mathbb{E}(\hat{r}(X) - r(X))^2 \geq -N \ln N + cN,$$

for some constant $c > 0$. By (3.6), we deduce that

$$\liminf_{n \rightarrow \infty} \frac{\ln \mathbb{E}(\hat{r}(X) - r(X))^2}{\sqrt{\ln n \ln_2 n}} \geq -\sqrt{\frac{\alpha}{2d}},$$

which, combined with Theorem 2.1, gives the result. \square

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

$$\hat{g}_{k,i}(x, y) = \frac{W_h^{\otimes k}(x - y)}{(\hat{\phi}_i + \rho)^{\otimes k}(x)}. \quad (3.7)$$

We also let for all $i = 1, \dots, n$:

$$d\hat{X}_i = dX_i - \hat{\phi}_i d\lambda, \quad d\bar{X}_i = dX - \hat{\phi}_i d\lambda, \quad (3.8)$$

$$d\tilde{X}_i = dX_i - \phi d\lambda \quad \text{and} \quad d\tilde{X} = dX - \phi d\lambda. \quad (3.9)$$

With this respect, we have:

$$\hat{I}_k = \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k^2} \hat{g}_{k,i}(x, y) \hat{X}_i^{\otimes k}(dy) \bar{X}_i^{\otimes k}(dx).$$

Proof of Lemma 3.1. For simplicity, we shall assume in this proof that $|Y| \leq 1$. Moreover, $C \geq 1$ denotes a constant whose value may change from line to line. With the help of notations (3.7)-(3.9), we let:

$$\begin{aligned}\hat{J}_1 &= \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k^2} \hat{g}_{k,i}(x,y) \tilde{X}_i^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x) \\ \hat{J}_2 &= \frac{1}{n} \sum_{i=1}^n Y_i \int_{\Delta_k^2} g_k(x,y) \tilde{X}_i^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x),\end{aligned}$$

where, for $x, y \in \mathbb{X}^k$:

$$g_k(x,y) = \frac{W_h^{\otimes k}(x-y)}{\varphi^{\otimes k}(x)}.$$

Then, since $|Y| \leq 1$, we get by Jensen's Inequality:

$$\begin{aligned}\mathbb{E}(\hat{I}_k - \hat{J}_1)^2 &\leq \mathbb{E} \left(\int_{\Delta_k^2} \hat{g}_{k,1}(x,y) [\tilde{X}_1^{\otimes k}(\mathrm{d}y) \tilde{X}_1^{\otimes k}(\mathrm{d}x) - \tilde{X}_1^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x)] \right)^2 \\ &\leq 2(R_{1k} + R_{2k}),\end{aligned}$$

where R_{1k} and R_{2k} are defined in Lemma 4.4. Hence,

$$\mathbb{E}(\hat{I}_k - \hat{J}_1)^2 \leq C^k (k!)^2 (1 + \beta_{2k}) \frac{b^2}{h^{dk}}. \quad (3.10)$$

Moreover, conditioning first by X_1, \dots, X_n then by X_2, \dots, X_n , we find with two successive applications of the isometry formula (1.2), that

$$\begin{aligned}\mathbb{E}(\hat{J}_1 - \hat{J}_2)^2 &\leq \mathbb{E} \left(\int_{\Delta_k^2} (\hat{g}_{k,1}(x,y) - g_k(x,y)) \tilde{X}_1^{\otimes k}(\mathrm{d}y) \tilde{X}^{\otimes k}(\mathrm{d}x) \right)^2 \\ &\leq (k!)^2 \int_{\mathbb{X}^{2k}} \mathbb{E}(\hat{g}_{k,1}(x,y) - g_k(x,y))^2 \varphi^{\otimes k}(x) \varphi^{\otimes k}(y) \mathrm{d}x \mathrm{d}y.\end{aligned}$$

Thus, using the notation of Lemma 4.3, we get:

$$\begin{aligned}\mathbb{E}(\hat{J}_1 - \hat{J}_2)^2 &\leq C^k (k!)^2 N_k \int_{\mathbb{X}^{2k}} (W_h^{\otimes k}(x-y))^2 \mathrm{d}x \mathrm{d}y \\ &\leq C^k (k!)^2 \frac{b^2 + \rho^2 + \beta_{2k}}{h^{dk}}\end{aligned} \quad (3.11)$$

Finally, formula (1.2) gives

$$\begin{aligned}\mathbb{E}(\hat{J}_2 - I_k(f_k))^2 &= \int_{\mathbb{X}^k} \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n Z_i(x) - f_k(x) \right)^2 \varphi^{\otimes k}(x) dx \\ &= \int_{\mathbb{X}^k} \left(\frac{1}{n} \text{var}(Z_1(x)) + (\mathbb{E}Z_1(x) - f_k(x))^2 \right) \varphi^{\otimes k}(x) dx,\end{aligned}\quad (3.12)$$

where for all $x \in \mathbb{X}^k$ and $i = 1, \dots, n$:

$$Z_i(x) = Y_i \int_{\Delta_k} g_k(x, y) \tilde{X}_i^{\otimes k}(dy).$$

By (1.3) and (1.2):

$$\begin{aligned}\mathbb{E}Z_1(x) &= \mathbb{E}r(X) \int_{\Delta_k} g_k(x, y) \tilde{X}^{\otimes k}(dy) = \int_{\mathbb{X}^k} f_k(y) g_k(x, y) \varphi^{\otimes k}(y) dy \\ &= \frac{1}{\varphi^{\otimes k}(x)} \int_{\mathbb{X}^k} f_k(y) W_h^{\otimes k}(x - y) \varphi^{\otimes k}(y) dy.\end{aligned}$$

Then, easy calculations prove that, since W has a compact support:

$$\int_{\mathbb{X}^k} (\mathbb{E}Z_1(x) - f_k(x))^2 \varphi^{\otimes k}(x) dx \leq C^k h. \quad (3.13)$$

Moreover, by (1.2):

$$\begin{aligned}\text{var}(Z_1(x)) &\leq \mathbb{E} \left(\int_{\Delta_k} g_k(x, y) X^{\otimes k}(dy) \right)^2 \\ &\leq k! \int_{\mathbb{X}^k} g_k^2(x, y) \varphi^{\otimes k}(y) dy \\ &\leq \frac{k!}{[\varphi^{\otimes k}(x)]^2} \int_{\mathbb{X}^k} [W_h^{\otimes k}(x - y)]^2 \varphi^{\otimes k}(y) dy.\end{aligned}$$

Hence,

$$\int_{\mathbb{X}^k} \text{var}(Z_1(x)) \varphi^{\otimes k}(x) dx \leq k! \frac{C^k}{h^{dk}}. \quad (3.14)$$

We can conclude with (3.12)-(3.14) that

$$\mathbb{E}(\hat{J}_2 - I_k(f_k))^2 \leq C^k \left(h + \frac{k!}{nh^{dk}} \right).$$

Finally, the lemma is then a consequence of (3.10), (3.11) and above, since b and ρ vanish. \square

4 Auxiliary results

4.1 Intensity estimation

Recall that

$$\beta_k = \rho^{-k} \exp\left(-\frac{nb^d \rho^2}{4\kappa}\right).$$

where $\kappa > 0$ is such that $\sup_{\mathbb{X}} \varphi \sup_{\mathbb{R}^{2d}} K \leq \kappa$.

Lemma 4.1. *Assume **H2** holds and $nb^{d+2} \geq 1$. Then, there exists $C \geq 1$ such that for all $k \geq 1$ and $x \in \mathbb{X}$:*

$$\begin{aligned} (i) \quad & \mathbb{E}|\hat{\varphi}_1(x) - \varphi(x)|^k \leq C^k b^k \\ (ii) \quad & \mathbb{E}\left|\frac{1}{\hat{\varphi}_1(x) + \rho} - \frac{1}{\varphi(x)}\right|^k \leq C^k (b^k + \rho^k + \beta_k). \end{aligned}$$

Proof. First we compute the bias of $\hat{\varphi}_1(x)$. For simplicity, we assume $d = 1$ only for the computation of the bias. By the very definition of $\hat{\varphi}_1(x)$ (see 2.4 and below), we have

$$\begin{aligned} \mathbb{E}\hat{\varphi}_1(x) &= \int_{-M}^M K_b(x, y) \varphi(y) dy \\ &= \int_{(x-M)/b}^{(x+M)/b} J(z) \varphi(x - bz) dz + \int_{(M+x)/b}^{(3M+x)/b} J(z) \varphi(bz - 2M - x) dz \\ &\quad + \int_{(M-x)/b}^{(3M-x)/b} J(z) \varphi(2M - x - bz) dz. \end{aligned}$$

Distinguishing the cases $x \in [-M, -M + b]$, $x \in [-M + b, M - b]$ and $x \in [M - b, M]$, it is an easy exercise to prove that under the Lipschitz condition **H2** on φ , we have

$$|\mathbb{E}\hat{\varphi}_1(x) - \varphi(x)| \leq 3bL_1 \int_{-1}^1 (|y| + 2)J(y)dy \leq Cb, \quad (4.1)$$

where, here and in the following, $C \geq 1$ is a constant that does not depend on x , k and n , and may change from line to line.

Our second task is to give an exponential inequality for the deviation probability of $\hat{\varphi}_1(x)$. Fix $\alpha > 0$ and, for simplicity, write $F_x(y) = K_b(x, y)$ for $y \in \mathbb{X}$. We have by independence, for all $s > 0$:

$$\begin{aligned} \mathbb{P}(\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x) \geq \alpha) &= \mathbb{P}\left(\exp\left(s \sum_{i=2}^n \int_{\mathbb{X}} F_x(dX_i - \varphi d\lambda)\right) \geq e^{\alpha sn}\right) \\ &\leq e^{-\alpha sn} \left(\mathbb{E} \exp\left(s \int_{\mathbb{X}} F_x(dX_1 - \varphi d\lambda)\right)\right)^{n-1}, \end{aligned}$$

using Markov Inequality. By Campbell Inequality (see [7]), we thus have

$$\mathbb{P}(\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x) \geq \alpha) \leq \exp\left(-\alpha sn + (n-1) \int_{\mathbb{X}} (e^{sF_x} - sF_x - 1) \varphi d\lambda\right). \quad (4.2)$$

But, if $K_+, \varphi_+ > 0$ are constants such that $K(x, y) \leq K_+$ and $\varphi(x) \leq \varphi_+$ for all $x, y \in \mathbb{X}$, we have:

$$\begin{aligned} \int_{\mathbb{X}} |e^{sF_x} - sF_x - 1| \varphi d\lambda &\leq \sum_{j \geq 2} \frac{s^j}{j!} \int_{\mathbb{X}} F_x^j \varphi d\lambda \\ &\leq \sum_{j \geq 2} \frac{s^j K_+^{j-1} \varphi_+}{j! b^{d(j-1)}} \\ &\leq \frac{\varphi_+ b^d}{K_+} \left(\exp\left(\frac{sK_+}{b^d}\right) - \frac{sK_+}{b^d} - 1 \right). \end{aligned}$$

Letting δ the function defined for all $t \geq 0$ by $\delta(t) = t - (1+t) \ln(1+t)$ and choosing s so that

$$s = \frac{b^d}{K_+} \ln\left(1 + \frac{\alpha}{\varphi_+}\right),$$

we have by (4.2) and above:

$$\begin{aligned} \mathbb{P}(\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x) \geq \alpha) &\leq \exp\left(\frac{nb^d \varphi_+}{K_+} \delta\left(\frac{\alpha}{\varphi_+}\right) - \frac{\alpha}{\varphi_+} + \frac{b^d \varphi_+}{K_+} \ln\left(1 + \frac{\alpha}{\varphi_+}\right)\right) \\ &\leq \left(1 + \frac{\alpha}{\varphi_+}\right) \exp\left(\frac{nb^d \varphi_+}{K_+} \delta\left(\frac{\alpha}{\varphi_+}\right)\right), \end{aligned}$$

for n large enough, since b vanishes. Considering $-F_x$ instead of F_x , we can conclude that

$$\mathbb{P}(|\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x)| \geq \alpha) \leq 2\left(1 + \frac{\alpha}{\varphi_+}\right) \exp\left(\frac{nb^d \varphi_+}{K_+} \delta\left(\frac{\alpha}{\varphi_+}\right)\right). \quad (4.3)$$

We are now in a position to establish (i) and (ii). Regarding (i), we have by (4.1):

$$\mathbb{E}|\hat{\varphi}_1(x) - \varphi(x)|^k \leq 2^k (C^k b^k + \mathbb{E}|\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x)|^k).$$

Moreover, writing

$$\mathbb{E}|\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x)|^k = \int_0^\infty \mathbb{P}(|\hat{\varphi}_1(x) - \mathbb{E}\hat{\varphi}_1(x)| \geq u^{1/k}) du$$

and observing that, since $\delta(t)$ is smaller than $-t^2/4$ or $-t/4$, depending on $t \leq 1$ or not, we deduce from (4.3) and an obvious decomposition of the above integral that :

$$\mathbb{E}|\hat{\phi}_1(x) - \mathbb{E}\hat{\phi}_1(x)|^k \leq \left(\frac{C}{\sqrt{nb^d}}\right)^k.$$

Putting all pieces together gives (i), since $nb^{d+2} \geq 1$.

Next we prove (ii). Note that, according to (4.1):

$$\left|\frac{1}{\mathbb{E}\hat{\phi}_1(x)} - \frac{1}{\varphi(x)}\right| \leq \frac{Cb}{\varphi_-(\varphi_- - Cb)}.$$

Since b vanishes as n tends to infinity, we thus have

$$\left|\frac{1}{\mathbb{E}\hat{\phi}_1(x)} - \frac{1}{\varphi(x)}\right| \leq Cb. \quad (4.4)$$

Let now $A = \{|\hat{\phi}_1(x) - \mathbb{E}\hat{\phi}_1(x)| \geq \rho\}$. Since $\inf_{\mathbb{X}} \varphi > 0$ by assumption, we have according to (4.1):

$$\begin{aligned} \mathbb{E}\left|\frac{1}{\hat{\phi}_1(x) + \rho} - \frac{1}{\mathbb{E}\hat{\phi}_1(x)}\right|^k &\leq \mathbb{E}\left|\frac{1}{\hat{\phi}_1(x) + \rho} - \frac{1}{\mathbb{E}\hat{\phi}_1(x)}\right|^k \mathbf{1}_{A^c} \\ &\quad + \mathbb{E}\left|\frac{1}{\hat{\phi}_1(x) + \rho} - \frac{1}{\mathbb{E}\hat{\phi}_1(x)}\right|^k \mathbf{1}_A \\ &\leq C^k (\rho^k + \rho^{-k} \mathbb{P}(A)) \\ &\leq C^k \left(\rho^k + \rho^{-k} \exp\left(-\frac{nb^d \rho^2}{4\kappa}\right)\right). \end{aligned}$$

Last inequality is a consequence of (4.3), $\rho \rightarrow 0$ as $n \rightarrow \infty$ and the fact that $\delta(t) \leq -t^2/4$ provided $t > 0$ is small enough. We can now conclude with the above inequality and (4.4). \square

4.2 Perturbated chaos

Lemma 4.2. *Let $g \in \mathbb{L}^2(\mu^{\otimes k})$ and $\psi \in \mathbb{L}^2(\lambda)$. If $d\nu = \psi d\lambda$, we have*

$$\begin{aligned} &\mathbb{E}\left(\int_{\Delta_k} g d[(X - \nu)^{\otimes k} - (X - \mu)^{\otimes k}]\right)^2 \\ &\leq \sum_{i=0}^{k-1} i! \binom{k}{i}^2 \|\varphi - \psi\|_2^{2(k-i)} \int_{\mathbb{X}^k} g(x)^2 \prod_{j=1}^i \varphi(x_j) \lambda^{\otimes k}(dx). \end{aligned}$$

Proof. For simplicity of the proof, we assume that g is a symmetric function. Otherwise, one only needs to consider its symmetrized version. First observe that by symmetry of g :

$$\int_{\Delta_k} g d(X - \nu)^{\otimes k} = \sum_{i=0}^k \binom{k}{i} \int_{\Delta_k} g d(X - \mu)^{\otimes i} d(\mu - \nu)^{\otimes k-i}.$$

Consequently,

$$\begin{aligned} D &= \int_{\Delta_k} g d(X - \nu)^{\otimes k} - \int_{\Delta_k} g d(X - \mu)^{\otimes k} \\ &= \sum_{i=0}^{k-1} \binom{k}{i} \int_{\Delta_k} g d(\mu - \nu)^{\otimes k-i} d(X - \mu)^{\otimes i}. \end{aligned}$$

Hence, letting for $i = 1, \dots, k$ and $x \in \mathbb{X}^i$,

$$\Gamma(x_1, \dots, x_i) = \{y \in \Delta_{k-i} : y_\ell \notin \{x_1, \dots, x_i\} \text{ for all } \ell = 1, \dots, k-i\},$$

we have:

$$D = \int_{\mathbb{X}^k} g(\varphi - \psi)^{\otimes k} d\lambda^{\otimes k} + \sum_{i=1}^{k-1} \binom{k}{i} \int_{\Delta_i} g_i d(X - \mu)^{\otimes i},$$

where for $(x_1, \dots, x_i) \in \Delta_i$,

$$g_i(x_1, \dots, x_i) = \int_{\Gamma(x_1, \dots, x_i)} g(x_1, \dots, x_k) \prod_{j=i+1}^k (\varphi - \psi)(x_j) \lambda(dx_j).$$

Observe that by Cauchy-Schwarz,

$$g_i(x_1, \dots, x_i)^2 \leq \|\varphi - \psi\|_2^{2(k-i)} \int_{\mathbb{X}^{k-i}} g(x_1, \dots, x_k) \lambda(dx_{i+1}) \cdots \lambda(dx_k).$$

Then, since each g_i is symmetric, we deduce from equations (1.1) and (1.2) that

$$\begin{aligned} \mathbb{E}D^2 &= \left(\int_{\mathbb{X}^k} g(\varphi - \psi)^{\otimes k} d\lambda^{\otimes k} \right)^2 + \sum_{i=1}^{k-1} i! \binom{k}{i}^2 \int_{\mathbb{X}^i} g_i^2 d\mu^{\otimes i} \\ &\leq \sum_{i=0}^{k-1} i! \binom{k}{i}^2 \|\varphi - \psi\|_2^{2(k-i)} \int_{\mathbb{X}^k} g(x)^2 \prod_{j=1}^i \varphi(x_j) \lambda^{\otimes k}(dx), \end{aligned}$$

hence the lemma. \square

4.3 Technical inequalities

Lemma 4.3. *Assume **H2** holds and $nb^{d+2} \geq 1$. There exists $C \geq 1$ such that for all $k, \ell, n \geq 1$:*

$$(i) \quad M_{k,\ell} = \sup_{x \in \mathbb{X}^k} \mathbb{E} \left(\frac{\|\hat{\phi}_1 - \varphi\|_2^\ell}{(\hat{\phi}_1 + \rho)^{\otimes k}(x)} \right)^2 \leq C^k (1 + \beta_{2k}) b^{2\ell};$$

$$(ii) \quad N_k = \sup_{x \in \mathbb{X}^k} \mathbb{E} \left| \frac{1}{(\hat{\phi}_1 + \rho)^{\otimes k}(x)} - \frac{1}{\varphi^{\otimes k}(x)} \right|^2 \leq C^k (b^2 + \rho^2 + \beta_{2k}).$$

Proof. We only prove (i). First observe that by Cauchy-Schwarz and since \mathbb{X} is bounded:

$$\begin{aligned} M_{k,\ell}^2 &\leq \mathbb{E} \|\hat{\phi}_1 - \varphi\|_2^{4\ell} \sup_{x \in \mathbb{X}^k} \mathbb{E} ((\hat{\phi}_1 + \rho)^{\otimes k}(x))^{-4} \\ &\leq C^k \sup_{x \in \mathbb{X}} \mathbb{E} |\hat{\phi}_1(x) - \varphi(x)|^{4\ell} \sup_{x \in \mathbb{X}^k} \mathbb{E} ((\hat{\phi}_1 + \rho)^{\otimes k}(x))^{-4}, \end{aligned}$$

where, here and in the following, C is a positive constant that does not depend on k, ℓ and n and may change from line to line. Hence by Lemma 4.1:

$$M_{k,\ell}^2 \leq C^k b^{4\ell} \sup_{x \in \mathbb{X}} \mathbb{E} ((\hat{\phi}_1 + \rho)^{\otimes k}(x))^{-4}. \quad (4.5)$$

Thus, we only need to consider the rightmost term. Fix $x \in \mathbb{X}^k$, and note that

$$\mathbb{E} ((\hat{\phi}_1 + \rho)^{\otimes k}(x))^{-4} \leq \frac{8}{(\varphi^{\otimes k}(x))^4} + 8 \mathbb{E} \left| \frac{1}{(\hat{\phi}_1 + \rho)^{\otimes k}(x)} - \frac{1}{\varphi^{\otimes k}(x)} \right|^4. \quad (4.6)$$

The task is to bound the term

$$A = \mathbb{E} \left| \frac{1}{(\hat{\phi}_1 + \rho)^{\otimes k}(x)} - \frac{1}{\varphi^{\otimes k}(x)} \right|^4.$$

We shall make use of the following inequality:

$$\left| \prod_{i=1}^k a_i - \prod_{i=1}^k b_i \right|^4 \leq 16^k \sum_{\emptyset \neq I \subset \{1, \dots, k\}} \prod_{i \in I} |a_i - b_i|^4 \prod_{i \notin I} b_i^4,$$

where the a_i 's and the b_i 's are positive real numbers. Since φ is bounded below by a positive constant:

$$\begin{aligned} A &\leq C^k \sum_{\emptyset \neq I \subset \{1, \dots, k\}} \mathbb{E} \prod_{i \in I} \left| \frac{1}{\hat{\varphi}_1(x_i) + \rho} - \frac{1}{\varphi(x_i)} \right|^4 \\ &\leq C^k \sum_{\emptyset \neq I \subset \{1, \dots, k\}} \prod_{i \in I} \left(\mathbb{E} \left| \frac{1}{\hat{\varphi}_1(x_i) + \rho} - \frac{1}{\varphi(x_i)} \right|^{4|I|} \right)^{1/|I|}, \end{aligned}$$

according to Hölder Inequality, and where $|I|$ is the cardinality of the set I . Thus, by Lemma 4.1:

$$\begin{aligned} A &\leq C^k \sum_{j=1}^k \binom{k}{j} \sup_{x \in \mathbb{X}} \mathbb{E} \left| \frac{1}{\hat{\varphi}_1(x) + \rho} - \frac{1}{\varphi(x)} \right|^{4j} \\ &\leq C^k \sum_{j=1}^k \binom{k}{j} \left(b^{4j} + \rho^{4j} + \rho^{-4j} \exp\left(-\frac{nb^d \rho^2}{4\kappa}\right) \right) \\ &\leq C^k \left(1 + \rho^{-4k} \exp\left(-\frac{nb^d \rho^2}{4\kappa}\right) \right). \end{aligned}$$

because b and ρ vanishes as $n \rightarrow \infty$. Assertion (i) is then a straightforward consequence of inequalities (4.5) and (4.6). \square

Before statement of next lemma, we recall the notations (3.7)-(3.9).

Lemma 4.4. *Assume **H2** and **H3** hold, and $nb^{d+2} \geq 1$. Then, there exists a constant $C \geq 1$ such that for all $k, n \geq 1$, both quantities above:*

$$\begin{aligned} R_{1k} &= \mathbb{E} \left(\int_{\Delta_k^2} \hat{g}_{k,1}(x, y) \hat{X}_1^{\otimes k}(\mathrm{d}y) (\bar{X}_1^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)) \right)^2 \\ R_{2k} &= \mathbb{E} \left(\int_{\Delta_k^2} \hat{g}_{k,1}(x, y) (\hat{X}_1^{\otimes k}(\mathrm{d}y) - \tilde{X}_1^{\otimes k}(\mathrm{d}y)) \tilde{X}^{\otimes k}(\mathrm{d}x) \right)^2 \end{aligned}$$

are bounded by

$$C^k (k!)^2 (1 + \beta_{2k}) \frac{b^2}{h^{dk}}.$$

Proof. We only prove the bound for R_{1k} , other proof being similar. In the sequel, $C \geq 1$ is a constant that does not depend on n and k , and may change from line to

line. Writing (see notations (3.7)-(3.9)):

$$R_{1k} = \mathbb{E} \mathbb{E} \left[\left\{ \int_{\Delta_k} \left(\int_{\Delta_k} \hat{g}_{k,1}(x, y) \hat{X}_1^{\otimes k}(\mathrm{d}y) \right) (\bar{X}_1^{\otimes k}(\mathrm{d}x) - \tilde{X}^{\otimes k}(\mathrm{d}x)) \right\}^2 \middle| X_1, \dots, X_n \right],$$

we obtain with Lemma 4.2, using the independence of X and X_1, \dots, X_n and the fact that φ is a bounded function:

$$R_{1k} \leq C^k \sum_{i=0}^{k-1} i! \binom{k}{i}^2 \mathbb{E} \left[V^{k-i} \int_{\mathbb{X}^k} \left(\int_{\Delta_k} \hat{g}_{k,1}(x, y) \hat{X}_1^{\otimes k}(\mathrm{d}y) \right)^2 \lambda^{\otimes k}(\mathrm{d}x) \right], \quad (4.7)$$

where $V = \|\hat{\varphi}_1 - \varphi\|_2^2$. Now fix $x \in \mathbb{X}^k$ and $i = 0, \dots, k-1$. We have

$$\mathbb{E} V^{k-i} \left(\int_{\Delta_k} \hat{g}_{k,1}(x, y) \hat{X}_1^{\otimes k}(\mathrm{d}y) \right)^2 \leq 2(A_1 + A_2), \quad (4.8)$$

where

$$\begin{aligned} A_1 &= \mathbb{E} V^{k-i} \left(\int_{\Delta_k} \hat{g}_{k,1}(x, y) \tilde{X}_1^{\otimes k}(\mathrm{d}y) \right)^2, \\ \text{and } A_2 &= \mathbb{E} V^{k-i} \left(\int_{\Delta_k} \hat{g}_{k,1}(x, y) (\hat{X}_1^{\otimes k}(\mathrm{d}y) - \tilde{X}_1^{\otimes k}(\mathrm{d}y)) \right)^2. \end{aligned}$$

We proceed to bound A_2 . As before, we apply Lemma 4.2, but conditionally on X_2, \dots, X_n . Hence, since φ , \mathbb{X} and W are bounded:

$$\begin{aligned} A_2 &\leq C^k \sum_{j=0}^{k-1} j! \binom{k}{j}^2 \mathbb{E} V^{2k-i-j} \int_{\mathbb{X}^k} \hat{g}_{k,1}^2(x, y) \lambda^{\otimes k}(\mathrm{d}y) \\ &\leq \frac{C^k}{h^{dk}} \sum_{j=0}^{k-1} j! \binom{k}{j}^2 \mathbb{E} \frac{V^{2k-i-j}}{[(\hat{\varphi}_1 + \rho)^{\otimes k}(x)]^2}. \end{aligned}$$

Consequently, by Lemma 4.3:

$$A_2 \leq \frac{C^k}{h^{dk}} (1 + \beta_{2k}) \sum_{j=0}^{k-1} j! \binom{k}{j}^2 b^{2(k-i-j)}. \quad (4.9)$$

In a similar fashion, we get by conditioning and (1.2):

$$\begin{aligned} A_1 &\leq \frac{C^k}{h^{dk}} k! \mathbb{E} V^{k-i} \int_{\mathbb{X}^k} \hat{g}_{k,1}^2(x, y) \lambda^{\otimes k}(\mathrm{d}y) \\ &\leq C^k k! (1 + \beta_{2k}) b^{2(k-i)}. \end{aligned}$$

Thus, by (4.7)-(4.9) and above:

$$\begin{aligned} R_{1k} &\leq \frac{C^k}{h^{dk}} (1 + \beta_{2k}) \sum_{i=0}^{k-1} i! \binom{k}{i}^2 \left(k! b^{2(k-i)} + \sum_{j=0}^{k-1} j! \binom{k}{j}^2 b^{2(2k-i-j)} \right) \\ &\leq C^k (k!)^2 (1 + \beta_{2k}) \frac{b^2}{h^{dk}}, \end{aligned}$$

hence the lemma. \square

References

- [1] Biau, G., Cérou, F. and Guyader, A. (2010). Rates of convergence of the functional k -nearest neighbor estimate, *IEEE Trans. Inf. Th.*, pp. 2034-2040.
- [2] Biau, G., Cadre, B. and Paris, Q. (2013). Cox process learning. Preprint.
- [3] Baïllo, A., Cuesta-Alberto, J. and Cuevas, A. (2011). Supervised Classification for a Family of Gaussian Functional Models, *Scand. J. Statist.*, pp. 480-498.
- [4] Cadre, B. (2013). Supervised classification of diffusion paths, *Math. Methods Statist.*, pp. 213-225.
- [5] Gyoöfi, L., Kohler, M., Krzyżak, A. and Walk, H. (2002). *A distribution-Free Theory of Nonparametric Regression*, Springer-Verlag, New-York.
- [6] Itô, K. (1956). Spectral type of the shift transformation of differential processes with stationary increments, *trans. Amer. Math. Soc.*, pp. 253-263.
- [7] Kingman, J.F.C. (1993). *Poisson Processes*, Oxford Studies in Probability.
- [8] Last, G. and Penrose, M.D. (2011). Poisson process Fock space representation, chaos expansion and covariance inequalities, *Probab. Theory Relat. Fields*, pp. 663-690.
- [9] Mecke, J. (1967). Stationaire zufällige Maße auf lokalkompakten abelschen Gruppen, *Z. Wahrsch. verw. Geb.*, pp. 36-58.
- [10] Nualart, D. and Vives, J. (1990). Anticipative calculus for the Poisson process based on the Fock space, *Séminaire de Probabilités XXIV*, Lectures Notes in Mathematics, pp. 154-165.
- [11] Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*, Springer-Verlag, New-York.
- [12] Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, Springer-Verlag, New-York.
- [13] Wiener, N. (1938). The homogeneous chaos, *AM. J. Math.*, pp. 897-936.