# NONPARAMETRIC REGRESSION ESTIMATION ONTO A POISSON POINT PROCESS COVARIATE\*

# BENOÎT CADRE<sup>1</sup> AND LIONEL TRUQUET<sup>2</sup>

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  $\ln \mathbb{E}(\hat{r}(X) - r(X))^2$ .

Mathematics Subject Classification. 62G05, 62G08.

Received March 13, 2014. Revised September 18, 2014.

## 1. INTRODUCTION

#### 1.1. Functional regression estimation

Let  $\mathscr{S}$  be a measurable space, and let the data  $(X_1, Y_1), \ldots, (X_n, Y_n)$  be independent  $\mathscr{S} \times \mathbb{R}$ -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  $\mathscr{S} = \mathbb{R}^d$ . We refer the reader to the book by Gyořfi *et al.* [7] 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,  $\mathscr{S}$  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 [13] provides a presentation of this area.

Keywords and phrases. Regression estimation, Poisson point process, Wiener-Itô decomposition, rates of convergence.

<sup>\*</sup> The authors wish to thank the referee and the Associate Editor for valuable comments and insightful suggestions.

<sup>&</sup>lt;sup>1</sup> IRMAR, ENS Rennes, CNRS, UEB, Campus de Ker Lann, Avenue Robert Schuman, 35170 Bruz, France. cadre@ens-rennes.fr

<sup>&</sup>lt;sup>2</sup> IRMAR, Ensai, CNRS, UEB, Campus de Ker Lann, Avenue Robert Schuman, 35170 Bruz, France. truquet@ensai.fr

#### B. CADRE AND L. TRUQUET

In the infinite-dimensional setting, the regression problem is faced with new challenges which requires changing methodology (see [2]). 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.* [5], Biau *et al.* [4], Cadre [6].

Among the classical models from the theory of time-dependent stochastic processes, the case of Poisson processes is of great interest. Here,  $\mathscr{S}$  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  $\mathscr{S}$  is identified to the so-called Poisson space over some measurable space  $\mathbb{X}$ , *i.e.* 

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

where  $\delta_x$  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 [9]) on a domain  $\mathbb{X} \subset \mathbb{R}^d$  with mean measure  $\mu$ , where  $\mu$  is a  $\sigma$ -finite measure on the Borel  $\sigma$ -field  $\mathscr{X}$  of  $\mathbb{X}$ . As seen before, this means X belongs to the space of integer-valued  $\sigma$ -finite measures on  $\mathbb{X}$  and satisfies:

- for any  $A \in \mathscr{X}$ , 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  $A_1, \ldots, A_\ell \in \mathcal{X}, X_{A_1}, \ldots, X_{A_\ell}$  are independent random variables.

In the case  $\mathbb{X} = \mathbb{R}$ , Itô's famous chaos expansion (see [8, 15]) 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 [12], and more recently by Last and Penrose [10].

Now recall some basic facts about chaotic decomposition in the Poisson space. Fix  $k \ge 1$ . 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  $\Delta_k = \{x \in \mathbb{X}^k : x_i \neq x_j \text{ for all } i \neq j\}$ . 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

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

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

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

the sum being taken over all permutations  $\sigma = (\sigma(1), \ldots, \sigma(k))$  of  $\{1, \ldots, k\}$ , and similarly for  $\bar{h}$ . In particular, note that  $I_k(g)$  is a square integrable random variable. In Nualart and Vives ([12], 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

$$r(X) = \mathbb{E}Y + \sum_{k \ge 1} \frac{1}{k!} I_k(f_k),$$
(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  $\mathbb{L}^2_{\text{sym}}(\mu^{\otimes k})$  – the subset of symmetric functions in  $\mathbb{L}^2(\mu^{\otimes k})$  –, and the decomposition is defined in an unique way. Last and Penrose [10] 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, and letting  $\ln_2 n = \ln(\ln n)$ :

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

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

In this paper, we focus on Poisson point process, since it models many interesting situations that naturally raise in statistics. However, we believe that our work may be generalized, for example towards the direction of Lévy processes for which a chaotic decomposition property also holds, see [1].

## 2. Regression estimate

#### 2.1. Model and heuristic of the estimate

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

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

Heuristics. In view of a short presentation of the heuristic of the estimate of r, we assume for simplicity that  $\varphi$  is a known positive function. Suppose also that the intensity  $\varphi$  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  $x \in \mathbb{X}$ :

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

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

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

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

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

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

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

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

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

By (1.1), the kth chaos  $I_k(f_k)$  is thus estimated by

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

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

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

where 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  $\varphi$  is known. Thus we shall first present a nonparametric estimator of  $\varphi$ , then we shall adapt the previous idea to this context.

## 2.2. Construction of the estimator

Estimation of  $\varphi$ . Construction of the nonparametric estimate of  $\varphi$  is based on Mecke's Formula (see [11]), which states in particular that

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

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

$$\hat{\varphi}(x) = \frac{1}{n} \sum_{j=1}^{n} \int_{\mathbb{X}} K_b(x, x') X_j(\mathrm{d}x'), \ x \in \mathbb{X},$$

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

$$K_b(x,u) = \prod_{j=1}^d \left( J_b(x_j - u_j) + J_b(2M + x_j + u_j) + J_b(2M - x_j - u_j) \right).$$
(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  $b \to 0$  as  $n \to \infty$ :

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

This particular construction is classical in order to avoid bias on the boundary of X, a classical drawback of usual convolution kernel estimators, for example when the intensity is positive on the boundary of X. In this case, we need to correct the bias using one-sided kernel near the boundary (see p. 30 in the book by Silverman [14]).

Then, we define for all  $i = 1, \ldots, n$  the leave-one-out nonparametric estimator  $\hat{\varphi}_i$  of  $\varphi$  by

$$\hat{\varphi}_i(x) = \hat{\varphi}(x) - \frac{1}{n} \int_{\mathbb{X}} K_b(x, x') X_i(\mathrm{d}x')$$
$$= \frac{1}{n} \sum_{j=1, j \neq i}^n \int_{\mathbb{X}} K_b(x, x') X_j(\mathrm{d}x'),$$

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 kth chaos  $I_k(f_k)$  is  $\hat{I}_k$ , such that

$$\hat{I}_{k} = \frac{1}{n} \sum_{i=1}^{n} Y_{i} \int_{\Delta_{k}} \left[ \int_{\Delta_{k}} \frac{W_{h}^{\otimes k}(x-\cdot)}{(\hat{\varphi}_{i}+\rho)^{\otimes k}(x)} (\mathrm{d}X_{i}-\hat{\varphi}_{i}\mathrm{d}\lambda)^{\otimes k} \right] (\mathrm{d}X-\hat{\varphi}_{i}\mathrm{d}\lambda)^{\otimes k} (x),$$

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

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

where N(n) tends to infinity.

### 2.3. Result

In the sequel,  $\|.\|$  is the euclidean norm on any  $\mathbb{R}^p$ . We assume that (X, Y) is independent from the sample  $(X_1, Y_1), \ldots, (X_n, Y_n)$ . 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  $x, y \in \mathbb{X}$ :

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

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

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

Recall that h = h(n), b = b(n) and  $\rho = \rho(n)$  are vanishing sequences of positive numbers. Moreover, N(n) tends to infinity; for simplicity, we write N instead of N(n).

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. H1 has been introduced for technical matters; however, similar results can be proved under tail assumptions on Y. Observe moreover that, since X is bounded, H3 holds when, for instance, the pair (X, Y) is such that each  $f_k$  can be written as  $f_k(x) = \theta^{\otimes k}(x)$  for all  $x \in \mathbb{X}^k$ , where  $\theta$  is a bounded Lipschitz function on X.

Finally, 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  $u_N/(N \ln N) \to 0$  as  $n \to \infty$ :

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

Moreover, since

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

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

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

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

In particular, for the following choices:

$$N = \left[\sqrt{\frac{2\alpha \ln n}{d \ln_2 n}}\right], \ h = e^{-N \ln N - u_N}, \tag{2.6}$$

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

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

Theorem 2.1 entails that, for all  $\omega < \alpha/(2d)$  and n large enough, the mean squared error between  $\hat{r}(X)$  and r(X) is at least

$$\exp\left(-\sqrt{\omega\ln n\ln_2 n}\right).$$

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

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

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 [7]), but faster than the rates obtained by Biau *et al.* [3] 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.* [3] utilize a k-nearest neighbor estimator whose construction does not depend on the law of X.

For fixed k, estimation of  $f_k$  more or less requires to make use of usual tools in nonparametric estimation in dimension dk. In particular, optimal estimator of  $f_k$  should be obtain for this specific problem. However, the mean squared error in the estimation of r(X) not only depends on the risks in the estimation of the  $f_k$ 's, but also on the remainder term of the chaotic decomposition which plays a key role. Thus, we proceed to a simultaneous optimization of N and the risks of the estimates of  $f_1, \ldots, f_N$ , leading to non-standard choices for the bandwidths. Nevertheless, due to the lack of information on the minimax rate of convergence in this infinite-dimensional setting, we can not guarantee that this procedure leads to an optimal estimate for r(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  $k \ge k_0$  have a null  $\mathbb{L}^2_{\text{sym}}(\mu^{\otimes k})$ -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  $u_N/N \to \infty$ , we have

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

To the best of our knowledge, this is the first exact (logarithmic) 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. Proofs 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  $\rho, b$  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  $nb^{d+2} \ge 1$ . Then, there exists  $C \ge 1$  such that for all  $k \ge 1$ :

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

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

$$\mathbb{E}\left(\sum_{k=1}^{N} \frac{1}{k!} \left(\hat{I}_{k} - I_{k}(f_{k})\right)\right)^{2} \leq C^{N} \sum_{k=1}^{N} \left(\frac{b^{2} + \beta_{2k}}{h^{dk}} + \frac{h}{(k!)^{2}} + \frac{1}{k!nh^{dk}}\right) \\ \leq C^{N} \left(\frac{b^{2} + \beta_{2N}}{h^{dN}} + h + \frac{1}{nh^{dN}}\right).$$
(3.1)

Moreover, according to Theorem 4.2 in Last and Penrose [10],

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

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

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

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

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

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:

$$\beta_{2N} \le \rho^{-2N} \mathrm{e}^{3N \ln \rho} \le \rho. \tag{3.3}$$

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

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

hence the first part of the theorem.

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

$$C^{N}h = e^{N\ln C - N\ln N - u_{N}} \le e^{-N\ln N + N\ln C}.$$
(3.5)

Furthemore, according to the Stirling's formula,

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

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

$$\ln \mathbb{E} \left( \hat{r}(X) - r(X) \right)^2 \leq \ln \left( e^{-N \ln N + N \ln C} + e^{-N \ln N + N(1 + \ln C)} \right)$$
$$\leq -N \ln N + \ln \left( e^{N \ln C} + e^{N(1 + \ln C)} \right).$$

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

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

Putting all pieces together yield

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

which is the desired result.

**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):

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

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

$$\mathbb{E}\left(\sum_{k=1}^{N}\frac{1}{k!}\left(\hat{I}_{k}-I_{k}(f_{k})\right)\right)^{2} \leq C^{N}h = e^{-N\ln N + N\ln C - u_{N}}.$$

Moreover, by the Stirling's formula:

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

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

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

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

$$\ln \mathbb{E} \left( \hat{r}(X) - r(X) \right)^2 \ge -N \ln N + cN$$

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

$$\liminf_{n \to \infty} \frac{\ln \mathbb{E} \left( \hat{r}(X) - r(X) \right)^2}{\sqrt{\ln n \ln_2 n}} \ge -\sqrt{\frac{\alpha}{2d}}$$

~

which, combined with Theorem 2.1, gives the result.

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

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

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

$$dX_i = dX_i - \hat{\varphi}_i d\lambda, \ dX_i = dX - \hat{\varphi}_i d\lambda,$$
(3.8)

$$dX_i = dX_i - \varphi d\lambda \text{ and } dX = dX - \varphi d\lambda.$$
 (3.9)

With this respect, we have:

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

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

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

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:

$$\mathbb{E}\left(\hat{I}_{k}-\hat{J}_{1}\right)^{2} \leq \mathbb{E}\left(\int_{\Delta_{k}^{2}}\hat{g}_{k,1}(x,y)\left[\hat{X}_{1}^{\otimes k}(\mathrm{d}y)\bar{X}_{1}^{\otimes k}(\mathrm{d}x)-\tilde{X}_{1}^{\otimes k}(\mathrm{d}y)\tilde{X}^{\otimes k}(\mathrm{d}x)\right]\right)^{2} \\ \leq 2(R_{1k}+R_{2k}),$$

259

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

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

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

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

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

$$\mathbb{E}\left(\hat{J}_{1}-\hat{J}_{2}\right)^{2} \leq C^{k}(k!)^{2}N_{k}\int_{\mathbb{X}^{2k}}\left(W_{h}^{\otimes k}(x-y)\right)^{2}\mathrm{d}x\mathrm{d}y$$
$$\leq C^{k}(k!)^{2}\frac{b^{2}+\rho^{2}+\beta_{2k}}{h^{dk}}.$$
(3.11)

Finally, formula (1.2) gives

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

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

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

By (1.3) and (1.2):

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

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

$$\int_{\mathbb{X}^k} \left( \mathbb{E} Z_1(x) - f_k(x) \right)^2 \varphi^{\otimes k}(x) \mathrm{d}x \le C^k h.$$
(3.13)

Moreover, by (1.2):

$$\operatorname{var} \left( Z_1(x) \right) \leq \mathbb{E} \left( \int_{\Delta_k} g_k(x, y) X^{\otimes k}(\mathrm{d}y) \right)^2$$
$$\leq k! \int_{\mathbb{X}^k} g_k^2(x, y) \varphi^{\otimes k}(y) \mathrm{d}y$$
$$\leq \frac{k!}{[\varphi^{\otimes k}(x)]^2} \int_{\mathbb{X}^k} \left[ W_h^{\otimes k}(x-y) \right]^2 \varphi^{\otimes k}(y) \mathrm{d}y.$$

Hence,

$$\int_{\mathbb{X}^k} \operatorname{var}\left(Z_1(x)\right) \varphi^{\otimes k}(x) \mathrm{d}x \le k! \frac{C^k}{h^{dk}}$$
(3.14)

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

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

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

# 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  $nb^{d+2} \ge 1$ . Then, there exists  $C \ge 1$  such that for all  $k \ge 1$  and  $x \in X$ :

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

*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

$$\begin{split} \mathbb{E}\hat{\varphi}_{1}(x) &= \int_{-M}^{M} K_{b}(x,y)\varphi(y)\mathrm{d}y\\ &= \int_{(x-M)/b}^{(x+M)/b} J(z)\varphi(x-bz)\mathrm{d}z + \int_{(M+x)/b}^{(3M+x)/b} J(z)\varphi(bz-2M-x)\mathrm{d}z\\ &+ \int_{(M-x)/b}^{(3M-x)/b} J(z)\varphi(2M-x-bz)\mathrm{d}z. \end{split}$$

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  $\varphi$ , we have

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

where, here and in the following,  $C \ge 1$  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:

$$\mathbb{P}\left(\hat{\varphi}_{1}(x) - \mathbb{E}\hat{\varphi}_{1}(x) \geq \alpha\right) = \mathbb{P}\left(\exp\left(s\sum_{i=2}^{n}\int_{\mathbb{X}}F_{x}(\mathrm{d}X_{i} - \varphi\mathrm{d}\lambda)\right) \geq \mathrm{e}^{\alpha s n}\right)$$
$$\leq \mathrm{e}^{-\alpha s n}\left(\mathbb{E}\exp\left(s\int_{\mathbb{X}}F_{x}(\mathrm{d}X_{1} - \varphi\mathrm{d}\lambda)\right)\right)^{n-1},$$

261

using Markov's inequality. By Campbell's inequality (see [9]), we thus have

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

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

Letting  $\delta$  the function defined for all  $t \ge 0$  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:

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

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

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

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

$$\mathbb{E}\left|\hat{\varphi}_{1}(x)-\varphi(x)\right|^{k} \leq 2^{k} \left(C^{k}b^{k}+\mathbb{E}\left|\hat{\varphi}_{1}(x)-\mathbb{E}\hat{\varphi}_{1}(x)\right|^{k}\right)$$

Moreover, writing

$$\mathbb{E} \left| \hat{\varphi}_1(x) - \mathbb{E} \hat{\varphi}_1(x) \right|^k = \int_0^\infty \mathbb{P} \left( \left| \hat{\varphi}_1(x) - \mathbb{E} \hat{\varphi}_1(x) \right| \ge u^{1/k} \right) \mathrm{d} u$$

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 :

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

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

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

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

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

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

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

$$\mathbb{E} \left| \frac{1}{\hat{\varphi}_{1}(x) + \rho} - \frac{1}{\mathbb{E}\hat{\varphi}(x)} \right|^{k} \leq \mathbb{E} \left| \frac{1}{\hat{\varphi}_{1}(x) + \rho} - \frac{1}{\mathbb{E}\hat{\varphi}(x)} \right|^{k} \mathbf{1}_{A^{c}} \\ + \mathbb{E} \left| \frac{1}{\hat{\varphi}_{1}(x) + \rho} - \frac{1}{\mathbb{E}\hat{\varphi}(x)} \right|^{k} \mathbf{1}_{A} \\ \leq C^{k} \left( \rho^{k} + \rho^{-k} \mathbb{P}(A) \right) \\ \leq C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^{d}\rho^{2}}{4\kappa} \right) \right) + C^{k} \left( \rho^{k} + \rho^{-k} \exp \left( -\frac{nb^$$

Last inequality is a consequence of (4.3),  $\rho \to 0$  as  $n \to \infty$  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).

## 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

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

*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:

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

Consequently,

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

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

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

we have:

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

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

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

Observe that by Cauchy–Schwarz,

$$g_i(x_1,\ldots,x_i)^2 \le \|\varphi-\psi\|_2^{2(k-i)} \int_{\mathbb{X}^{k-i}} g(x_1,\ldots,x_k)\lambda(\mathrm{d} x_{i+1})\ldots\lambda(\mathrm{d} x_k)$$

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

$$\mathbb{E}D^{2} = \left(\int_{\mathbb{X}^{k}} g(\varphi - \psi)^{\otimes k} \mathrm{d}\lambda^{\otimes k}\right)^{2} + \sum_{i=1}^{k-1} i! \binom{k}{i}^{2} \int_{\mathbb{X}^{i}} g_{i}^{2} \mathrm{d}\mu^{\otimes i}$$
$$\leq \sum_{i=0}^{k-1} i! \binom{k}{i}^{2} \|\varphi - \psi\|_{2}^{2(k-i)} \int_{\mathbb{X}^{k}} g(x)^{2} \prod_{j=1}^{i} \varphi(x_{j})\lambda^{\otimes k}(\mathrm{d}x),$$

hence the lemma.

# 4.3. Technical inequalities

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

(i) 
$$\begin{split} M_{k,\ell} &= \sup_{x \in \mathbb{X}^k} \mathbb{E} \left( \frac{\|\hat{\varphi}_1 - \varphi\|_2^\ell}{(\hat{\varphi}_1 + \rho)^{\otimes k}(x)} \right)^2 \leq C^k (1 + \beta_{2k}) b^{2\ell}; \\ (\text{ii}) \quad N_k &= \sup_{x \in \mathbb{X}^k} \mathbb{E} \Big| \frac{1}{(\hat{\varphi}_1 + \rho)^{\otimes k}(x)} - \frac{1}{\varphi^{\otimes k}(x)} \Big|^2 \leq C^k \left( b^2 + \rho^2 + \beta_{2k} \right). \end{split}$$

*Proof.* We only prove (i). First observe that by Cauchy–Schwarz and since X is bounded:

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

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

$$M_{k,\ell}^2 \le C^k b^{4\ell} \sup_{x \in \mathbb{X}} \mathbb{E}\left( (\hat{\varphi}_1 + \rho)^{\otimes k}(x) \right)^{-4}.$$

$$(4.5)$$

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

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

The task is to bound the term

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

264

We shall make use of the following inequality:

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

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

$$A \leq C^{k} \sum_{\substack{\emptyset \neq I \subset \{1,\dots,k\}}} \mathbb{E} \prod_{i \in I} \left| \frac{1}{\hat{\varphi}_{1}(x_{i}) + \rho} - \frac{1}{\varphi(x_{i})} \right|^{4}$$
$$\leq C^{k} \sum_{\substack{\emptyset \neq I \subset \{1,\dots,k\}}} \prod_{i \in I} \left( \mathbb{E} \left| \frac{1}{\hat{\varphi}_{1}(x_{i}) + \rho} - \frac{1}{\varphi(x_{i})} \right|^{4|I|} \right)^{1/|I|},$$

according to Hölder's inequality, and where |I| is the cardinality of the set I. Thus, by Lemma 4.1:

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

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

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

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

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

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,  $C \ge 1$  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)):

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

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

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

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

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

where

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

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

$$A_{2} \leq C^{k} \sum_{j=0}^{k-1} j! {\binom{k}{j}}^{2} \mathbb{E} V^{2k-i-j} \int_{\mathbb{X}^{k}} \hat{g}_{k,1}^{2}(x,y) \lambda^{\otimes k}(\mathrm{d}y)$$
$$\leq \frac{C^{k}}{h^{dk}} \sum_{j=0}^{k-1} j! {\binom{k}{j}}^{2} \mathbb{E} \frac{V^{2k-i-j}}{\left[(\hat{\varphi}_{1}+\rho)^{\otimes k}(x)\right]^{2}} \cdot$$

Consequently, by Lemma 4.3:

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

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

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

Thus, by (4.7)–(4.9) and above:

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

hence the lemma.

## References

- [1] D. Applebaum, Universal Malliavin calculus in Fock and Lévy-Itô Spaces. Commun. Stoch. Anal. (2009) 119-141.
- [2] J.-M. Azais and J.-C. Fort, Remark on the finite-dimensional character of certain results of functional statistics. C.R. Acad. Sci. **351** (2013) 139–141.
- [3] G. Biau, F. Cérou and A. Guyader, Rates of convergence of the functional k-nearest neighbor estimate. IEEE Trans. Inf. Theory 56 (2010) 2034–2040.
- [4] G. Biau, B. Cadre and Q. Paris, Cox process functional learning. To appear in Stat. Int. Stoch. Processes (2015).
- [5] A. Baillo, J. Cuesta-Alberto and A. Cuevas, Supervised classification for a family of Gaussian functional models. Scand. J. Statist. 38 (2011) 480–498.
- [6] B. Cadre, Supervised classification of diffusion paths. Math. Methods Statist. 22 (2013) 213–225.

- [7] L. Gyoïfi, M. Kohler, A. Krzyżak and H. Walk, A distribution-Free Theory of Nonparametric Regression. Springer-Verlag, New-York (2002).
- [8] K. Itô, Spectral type of the shift transformation of differential processes with stationary increments. Trans. Amer. Math. Soc. 81 (1956) 253-263.
- [9] J.F.C. Kingman, Poisson Processes. In Oxf. Stud. Probab. Oxford Science publications, 1st ed. (1993).
- [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] J. Mecke, Stationaire zufällige Ma $\beta$ e auf lokalkompakten abelschen Gruppen. Z. Wahrsch. verw. Geb. 9 (1967) 36–58.
- [12] D. Nualart and J. Vives, Anticipative calculus for the Poisson process based on the Fock space. Séminaire de Probabilités XXIV. Lect. Notes Math. (1990) 154–165.
- [13] J.O. Ramsay and B.W. Silverman, Functional Data Analysis. Springer-Verlag, New-York (1997).
- [14] B.W. Silverman, Density Estimation for Statistics and Data Analysis. Springer-Verlag, New-York (1986).
- [15] N. Wiener, The homogeneous chaos. Am. J. Math. 60 (1938) 897–936.