Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 51 (3) (2022), 914 – 931
DOI : 10.15672/hujms.796694

Research Article

The k nearest neighbors local linear estimator of functional conditional density when there are missing data

Ibrahim M. Almanjahie*[1,2], Wafaa Mesfer Alahmari[1], Ali Laksaci[1,2]

[1] *Department of Mathematics, College of Science, King Khalid University, Abha 62529, Saudi Arabia*
[2] *Statistical Research and Studies Support Unit, King Khalid University, Abha 62529, Saudi Arabia*

Abstract

Our key aim is to propose effective estimators for the conditional probability density of a scalar response variable given a functional co-variable, where the response variable is considered to have missing data at random. Such estimators are constructed by combining the approaches of the local linear method and the kernel nearest neighborhood. The main feature of this estimation is the possibility to model the missing phenomena. Under less restrictive conditions, we show the strong consistency of the proposed estimators. To assess the efficacy of the developed estimators, empirical analysis as well as real data analyses are performed.

Mathematics Subject Classification (2020). 62G08, 62G10, 62G35, 62G07, 62G32, 62G30, 62H12

Keywords. functional data analysis, small ball probability, local linear method, k nearest neighbors, conditional density, almost complete convergence

1. Introduction

For nonparametric statistics, the conditional density function plays a significant role. It is used to characterize the relationship between two phenomena in various applied areas. In particular, it helps in controlling the inflation rate in economics, predicting the pollution peaks in environments or constructing the region predictive in the analysis of time series problem.

Historically, the first study considering the estimation method in conditional density dates back to [33]. He proved the almost surely convergence of the kernel estimator of the transition probability density in the Markov chain. The issue of the automatic determination of the smoothing parameters in the estimation method of conditional density has been considered by [35]. We refer to [23] for the L_p consistency of the kernel estimation method using the conditional probability density in which the observations meet the strong markovian property. Some authors have treated the estimation method in the conditional density as a preliminary model of the conditional mode (see, for instance, [12, 28, 30]).

*Corresponding Author.
Email addresses: imalmanjahi@kku.edu.sa (I.M. Almanjahie), wafamesfer@gmail.com (W.M. Alahmari), alilak@yahoo.fr (A. Laksaci)
Received: 18.09.2021; Accepted: 25.03.2022

The conditional density function has been introduced first in the functional statistics area by [17]. They proposed the functional version of the Nadraya-Watson estimator in the independent identically distributed case and then determined its almost complete consistency. Later, the L_2 consistency of their estimator was proved by [22]. The well-known asymptotic normality of the kernel estimator has been considered by some researchers. We cite, for instance, Ezzahrioui and Ould Said [16], who proved this asymptotic property under the strong mixing condition. Considering the spatial dependent case, Dabo-Niang et al. [13] have studied the almost complete consistency and the L_p convergence of the spatial version of the functional Nadraya-Watson estimator of the conditional density function.

The local linear method has become of interest to many researchers. Baillo and Grane [6] were the first to introduce this topic in the functional statistics area. Based on a regression operator, they studied the scalar variable given a Hilberthian regressor. In contrast to the multivariate case, several approaches to the local linear estimation exist in functional statistics. For example, Barrientos-Marin et al. [7] have constructed a fast version that can be used even if the predictor variable belongs to Banach space. Boj et al. [9] have considered another alternative estimator for the functional local linear regression. All these cited studies concern the regression operator. Demongeot et al. [14] considered the conditional density and determined the first results of the functional local linear estimation. In their constructed estimator and using the fast version of local linear modeling proposed by [7], the uniform almost complete was established. It is notable that the principal feature of the local linear method is the likelihood to reduce the bias term when using classical kernel approach. Motivated by this feature, Rachdi et al. [32] considered the conditional density, studied the bias term for the functional local linear estimator and have quantified the gain on the bias. More recently, Almanjahie et al. [2] have constructed an alternative estimator using the k nearest neighbor (kNN) smoothing and established the almost complete consistency of their estimator. In parallel, Ferraty et al. [18] stated the nonparametric modeling of functional data with missing responses at random. Based on the regression operator, they used independent identically distributed (iid) data and studied the kernel estimator when the functional regressor is completely observed and the scalar response variable is Missing at Random (MAR). Later, their results were generalized to an ergodic functional time series case by [25]. Considering the same data structure, the problem of conditional mode estimation was investigated by [26]. Recently, Benchiha and Kaid [8] considered the regression operator and used the local linear estimation method in the MAR data case to construct the almost complete consistency of their estimator.

In this paper, the estimation method of the local linear approach for the Conditional Density Probability (CDP) is considered but when the regressors are infinite-dimensional. We construct two new estimators of the CDP which are constructed by combining the approaches of local linear method to the kNN approach . The principal advantage is that the built estimators inherit the good statistical properties of the two approaches. In particular, the smoothing parameter of the kNN is locally adapted to the structure of the data which is extremely beneficial in Functional Data Analysis (FDA) when the convergence rate and the local behavior of the data are closely linked. However, because of the randomness of the bandwidth parameter, the statement of the estimator asymptotic properties is complex and needs some further mathematical devolvement. In this context, the almost complete consistency)[†] of the two estimators will be proved. These asymptotic properties will be treated under general conditions allows highlighting the bi-functional

[†]We say that the sequence of random variables $(\boldsymbol{\Delta}_n)$ converges a.co. to zero, if and only if

$$\forall \tau > 0, \sum_{n\geq 1} \mathbb{P}(|\boldsymbol{\Delta}_n| > \tau) < \infty.$$

Furthermore, we say that $\boldsymbol{\Delta}_n = O_{a.co.}(\boldsymbol{\delta}_n)$, if there exists $\tau_0 > 0$, such that $\sum_{n\geq 1} \mathbb{P}(|\boldsymbol{\Delta}_n| > \tau_0 \boldsymbol{\delta}_n) < \infty$.

dimensionality of both models and data. This study is illustrated based on real data examples and some Monte Carlo studies. Finally, let us stress that the statistical method for analyzing functional data is actually in a continuous devolvement. The reader interested by this topic may refer to the survey paper by [27] or the work by [3] for recent advances and references. Meanwhile, for a background in kNN smoothing and/or nonparametric modeling for functional MAR data, we refer to [4, 5, 10, 20, 31].

The outline of this paper is as follows. The presentation of the general framework of this paper is introduced in Section 2. The construction of the estimators is performed in Section 3. Also, we establish the asymptotic properties of the constructed estimators. We devote Section 4 to the computational analysis of the constructed estimators. Such analysis allows us to highlight the real impact of the proposed model in practice. Our proofs of the technical lemmas are detailed in the appendix section. Conclusion and prospects are stated in Section 5.

2. The incomplete functional data framework

For $i = 1, \dots, n$, let (X_i, Y_i) represent n pairs of independent random vectors which are drawn from $(X, Y) \in \mathcal{F} \times \mathbb{R}$. Here, $\mathcal{F}$ represents a functional space, eventually, finite dimensional, Hilbert (equipped with the norm $\|.\|$) or semi-metric space (equipped with a semi-metric d). From now, we take x in $\mathcal{F}$(resp. y in $\mathbb{R}$) and the neighborhoods of x and y are respectively $\mathcal{N}_x$ and $\mathcal{N}_y$.

Next, we assume the conditional probability distribution of $Y \mid X$ exists and is absolutely continuous with respect the Lebesgue measure on $\mathbb{R}$. Note that the CDP function of $Y \mid X = x$ is indicated by $\varphi(y|x)$. In addition, the functional variable X satisfies

$$\phi_x(r) := \mathbb{P}(X \in B(x, r)) > 0, \tag{2.1}$$

where $B(x, r)$ is the closed topology ball in $\mathcal{F}$ centered in x and r being the radius. It is defined by

$$B(x, r) = \{z \in \mathcal{F} : d(x, z) \leq r\}.$$

The function in Equation (2.1), is an invertible function, and if $0 < c < 1 < c^* < \infty$ exist, then we have the following condition:

$$\lim_{r \to 0} \frac{\phi_x(rc)}{\phi_x(r)} < 1 < \lim_{r \to 0} \frac{\phi_x(rc^*)}{\phi_x(r)}. \tag{2.2}$$

Considering the conditional density function and studying its local linear estimator is the main objective of the current paper. Define the cumulative conditional distribution function as

$$\Phi(x, y) = \mathbb{P}(Y \leq y | X = x).$$

Then, the CDP function is

$$\varphi(y|x) = \frac{\partial \Phi(x, y)}{\partial y}.$$

Note that the variable Y is the response variable. Therefore, the functional space of this nonparametric model is characterized using the following condition:

- If $(\mathcal{F}, \|.\|)$ is Hilbert space: There exist $\gamma > 0$, and $C > 0$ such that
$$\forall (y_1, y_2) \in \mathcal{N}_y \times \mathcal{N}_y, \quad |\varphi(y_1|x) - \varphi(y_2|x)| \leq C\left(|y_1 - y_2|^{\gamma}\right). \tag{2.3}$$
- If $(\mathcal{F}, d)$ is semi-metric space: There exist $C > 0$, $\gamma_1 > 0$, and $\gamma_2 > 0$ such that $\forall (y_1, y_2) \in \mathcal{N}_y \times \mathcal{N}_y$ and $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$
$$|\varphi(y_1|x_1) - \varphi(y_2|x_2)| \leq C\left(d^{\gamma_1}(x_1, x_2) + |y_1 - y_2|^{\gamma_2}\right). \tag{2.4}$$

The novelty of the current work depends on considering the general case that covers the MAR in Y. Such situation is modeled by introducing a variable δ which follows a Bernoulli distribution where $\delta = 1$ means the Y is observed, otherwise the $\delta = 0$. Considering the MAR means that for a given X, Y and δ are conditionally independent. Mathematically, we have

$$\mathbb{P}(\delta = 1|X, Y) = \mathbb{P}(\delta = 1|X) = P(X),$$

where the functional operator $P(\cdot)$ is unknown. If the explanatory variable X is known, then the function $P(\cdot)$ gives the probability to observe of Y. For the asymptotic study, we need the following regularity condition.

$$P(\cdot) \text{ is a continuous function on } \mathcal{N}_x \text{ and such that } P(z) > 0, \text{ for all } z \in \mathcal{N}_x. \tag{2.5}$$

Our objective is the construction and studying of the asymptotic property of two local linear estimators adapted to this incomplete functional data case. The first one is used for an structure while the second one is adapted for semi-metric functional space.

3. The kNN local linear estimation of the CDP

3.1. Hilbertian regressor case

In this part, $\mathcal{F}$ is assumed to be a separable Hilbert space on an orthonormal basis $(v_j)_{j\geq 1}$ and equipped with the norm $\|.\|$. As all local linear fitting, it is also assumed that the local approximation of $\varphi(y|x)$ is achieved by a linear function. In sense that, for all $x_0 \in \mathcal{N}_x$, we get

$$\varphi(y|x_0) = a_{y|x} + b_{y|x}(x_0 - x) + \rho_{y|x}(x_0 - x, x_0 - x) + o(\|x_0 - x\|^2). \tag{3.1}$$

Note that the linear $b_{y|x}$ and bilinear $\rho_{y|x}$, in Equation (3.1), are continuous operators. The former maps from $\mathcal{F}$ to $\mathbb{R}$ and the later maps from $\mathcal{F} \times \mathcal{F}$ to $\mathbb{R}$. As suggested in [1], the linearity property of $b_{y|x}$ together with the decomposition process of $(X_i - x)$ on $(v_j)_{j\geq 1}$ of $\mathcal{F}$ to a threshold J are used to estimate the coefficients $a_{y|x}$ and $b_{y|x}$. Precisely, As $J \to \infty$, we get the consistency of

$$\sum_{j=1}^{J} c_j v_j \to X - x$$

from Parseval's theorem. Note that the $(c_j)_j$ represents the $(X - x)$ coefficients in the $(v_j)_{j\geq 1}$ basis. Under this consideration, the Local Linear kNN Estimators (kNN-LLE) of $a_{y|x}$ and $b_{y|x}$ are determined from minimizing the criterion

$$\min_{a, b_1, \ldots, b_J \in \mathbb{R}} \sum_{i=1}^{n} \left(\ell_l^{-1} H(\ell_l^{-1}(y - Y_i) - a - \sum_{j=1}^{J} c_{ij} b_j \right)^2 \delta_i K\left(\frac{\|x - X_i\|}{h_k} \right),$$

where for $j = 1, \ldots, J$, we put $b_j = b_{y|x}(v_j)$ and K and H denote the kernels. The h_k and ℓ_l are the bandwidths, defined by

$$h_k = \min\{h \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^{n} \mathbb{1}_{B(x,h)}(X_i) = k\}$$

and

$$\ell_l = \min\{\ell \in \mathbb{R}^+, \text{ such that } \sum_{i=1}^{n} \mathbb{1}_{(y-\ell,\, y+\ell)}(Y_i) = l\}.$$

Here, the $\mathbb{1}_A$ refers to an indicator function on the set A. Note that the approach of kNN method was presented in FDA area by [10]. In spite of the significance of this methodology, practically speaking, the useful of kNN smoothing has not yet been completely investigated. Some contributions on this topic exist on the literature; see for instances, [20, 21, 24]. Finally, the Hilibethian version of kNN-LLE estimation of $\varphi(y|x)$ can be explicitly expressed by the matrix

$$\varphi_n(y|x) = \widehat{a}_{yx} = e_1^{J+1'}(Q'_B\mathbf{K}Q_B)^{-1}(Q'_B\mathbf{K}\mathbf{H}). \tag{3.2}$$

$e_1^{J+1'}$ is the transpose vector of the first canonical basis vector of $\mathbb{R}^{J+1}$ and Q_B is given by

$$Q_B = \begin{pmatrix} 1 & c_{11} & \dots & c_{1J} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & c_{n1} & \dots & c_{nJ} \end{pmatrix}.$$

We set

$$\mathbf{K} = diag(\delta_1 K(h_k^{-1}\|x - X_1\|), \dots, \delta_n K(h_k^{-1}\|x - X_n\|)),$$
$$\mathbf{H} = \ell_l^{-1}(H(\ell_l^{-1}(y - Y_1), \dots, H(\ell_l^{-1}(y - Y_n))).$$

The following conditions (A4-A6) are needed for establishing the almost complete consistency of the estimator $\varphi_n(y|x)$:

$$\text{The coefficients } (c_j)_j \text{ such that } \sum_{j=J+1} c_j^2 = O_{a.co}(J^{-v}) \quad \text{for certain} \quad v > 0. \tag{3.3}$$

The kernel K is a differentiable function which is supported within $(0, 1)$ such that

$$-\infty < C^* < K'(t) < C < 0 \quad \text{for} \quad 0 \le t \le 1, \quad C, C' > 0. \tag{3.4}$$

The kernel H is a continuous function has compact support and satisfies

$$\int H(t)dt = 1. \tag{3.5}$$

The number of nearest neighbors k and l satisfy that

$$\frac{n \log n}{lk} \to 0 \text{ as } n \to \infty. \tag{3.6}$$

Theorem 1. If the conditions (2.2), (2.3),(2.5)-(3.6) are satisfied, then

$$|\varphi(y|x) - \varphi_n(y|x)| = O(J^{-v}) + O\left(\frac{l}{n}\right)^{\gamma} + O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^2\right) + O_{a.co.}\left(\sqrt{\frac{n \log n}{lk}}\right), \tag{3.7}$$

as $\min(n, J) \to \infty$.

Proof: The proof is briefly presented. It is obtained by combining the ideas of [10] to those used by [31]. Throughout we set, for $\alpha_1, \alpha_2, \alpha_3 \in (0, 1)$ such that

$$\phi_x^{-1}\left(\frac{k}{\alpha_1 n}\right) \le C\phi_x^{-1}\left(\frac{\alpha_2 k}{n}\right),$$

by

$$h_k^r = \phi_x^{-1}(k/\alpha_1 n), \, h_k^l = \phi_x^{-1}(\alpha_2 k/n), \, \ell^l = \frac{l}{\alpha_3 n} \text{ and } \ell^r = \frac{\alpha_3 l}{n}. \tag{3.8}$$

Thus, Theorem 1 is a direct result of lemmas below.

Lemma 1. (See [11]) Under the conditions (2.2), (2.5), (3.3) and (3.4), we obtain, for all $j, j' = 1, \dots, J$,

$$S_{n,j',j} = O_{a.co.}(1),$$

where $S_{n,j',j} = \dfrac{1}{nh^2\phi_x(h)} \displaystyle\sum_{i=1}^{n} c_{ij'}c_{ij}\delta_i K(h^{-1}\|x - X_i\|)$. for $h = h_k^r$ or h_k^l.

Lemma 2. Using the Theorem 1conditions and for all $j = 1, \dots, J$, we get

$$\mathbb{E}\left[T_{n,j}\right] = O\left(\ell^{\gamma}\right) \quad \text{and} \quad \mathbb{E}\left[e_{n,j}\right] = O\left(h^2\right),$$

where $T_{n,j} = \frac{1}{nh\ell\phi_x(h)} \sum_{i=1}^{n} c_{ij}\delta_i K(h^{-1}\|x - X_i\|)(H(\ell^{-1}(y - Y_i) - \varphi(y|X_i))$ and $e_{n,j} = \frac{1}{nh\phi_x(h)} \sum_{i=1}^{n} c_{ij}\delta_i K(h^{-1}\|x - X_i\|)\rho_{xy}(X_i - x, X_i - x)$, for $h = h_k^r$ or h_k^l and $\ell = \ell_l^r$ or ℓ_l^l.

Lemma 3. Using the same conditions that used in Theorem 1 and for all $j = 1, \dots, J$, we get

$$T_{n,j} - \mathbb{E}\left[T_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right)$$

and

$$e_{n,j} - \mathbb{E}\left[e_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

Corollary 1. Under the same conditions that used in Theorem 1, we determine

$$e_1' S_n^{-1} T_n = O\left(\ell^{\gamma}\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right),$$

with

$$e_1' S_n^{-1} e_n = O\left(h^2\right) + O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right).$$

■

3.2. Functional regressor case

In this part, we work with more general structure by taking $\mathcal{F}$ to be a semi-metric space. This space is equipped with a semi-metric d and we adopt the fast version of the functional locally modeling that proposed by [7]. We work with the conditional density and considered the following local approximation. For all $x_0 \in \mathcal{N}_x$, and using

$$\varphi(y|x_0) = A(y|x) + B(y|x)\beta(x, x_0) + o(d(x_0, x)) \tag{3.9}$$

where $\beta(.,.)$ is a known function from $\mathcal{F}^2$ into $\mathbb{R}$ such that, $\forall x' \in \mathcal{F}$, $\beta(x', x') = 0$.

$$\forall x' \in \mathcal{F}, \; C_1 \left|\delta(x, x')\right| \leq \left|\beta(x, x')\right| \leq C_2 \left|\delta(x, x')\right|, \text{ where } C_1 > 0, \, C_2 > 0.$$

Then, the minimization of the following quantity

$$\min_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \left(\ell_l^{-1} H(\ell_l^{-1}(y - Y_i) - a - b\beta(X_i, x)\right)^2 \delta_i K(h_k^{-1} d(x, X_i)), \tag{3.10}$$

leads to the estimating of the scalars $A(y|x)$ and $B(y|x)$.

Note that the pair (x, y) is fixed and belongs to $\mathcal{F} \times \mathbb{R}$.Now, using the simple algebra as in the Hilbertian case, we prove that the minimizers of (3.10) are the solutions of

$$\begin{pmatrix} \widehat{A(y|x)} \\ \widehat{B(y|x)} \end{pmatrix} = (D'\mathbf{K}D)^{-1}(D'\mathbf{K}\mathbf{H}).$$

where

$$D = \begin{pmatrix} 1 & \beta(X_1, x) \\ \vdots & \vdots \\ 1 & \beta(X_n, x) \end{pmatrix}.$$

It follows that

$$\widehat{\varphi_n}(y|x) = \widehat{A}(y|x) = e_1^{2'}(D'\mathbf{K}D)^{-1}(D'\mathbf{K}\mathbf{H}).$$

Here $e_1^{2'}$ is the transpose vector of the first canonical basis vector of $\mathbb{R}^2$. Therefore it is explicitly defined by

$$\widehat{\varphi_n}(y|x) = \frac{\sum_{i,j=1}^n F_{ij}(x, h_k) H(\ell_l^{-1}(y - Y_j))}{\ell_l \sum_{i,j=1}^n F_{ij}(x, h_k)}, \tag{3.11}$$

where $\widehat{F}_{ij}(x, h_k) = \beta(X_i, x)\left(\beta(X_i, x) - \beta(X_j, x)\right) \delta_i \delta_j K(h_k^{-1} d(x, X_i)) K(h_k^{-1} d(x, X_j))$.

Constructing the almost complete consistency of the above estimator, in Equation (3.11), is not straightforward. To proceed, we need the following additional condition for achieving our goal.

$$\phi_x^{-1}\left(\frac{k}{n}\right) \int_{B(x,h_k)} \beta(u, x) dP(u) = o\left(\int_{B(x,h_k)} \beta^2(u, x)\, dP(u)\right) \tag{3.12}$$

where $dP(x)$ is the cumulative distribution of X.

Theorem 2. Based on the conditions (2.2), (2.4), (2.5) and (3.4)-(3.12), we have that:

$$|\widehat{\varphi_n}(y|x) - \varphi(y|x)| = O\left(\phi_x^{-1}\left(\frac{k}{n}\right)^{\gamma_1}\right) + O\left(\frac{k}{n}\right)^{\gamma_2} + O_{a.co.}\left(\sqrt{\frac{n \log n}{\ell k}}\right).$$

Proof: The proof is obtained by using a similar ideas to those used in Theorem 1. It will be presented in brief. It is based on the results of the following lemmas with $h =$ either h_k^r or h_k^l and $\ell =$ either ℓ_l^r or ℓ_l^l

Lemma 4. Based on the conditions of the Theorem 2, we get, for all $j = 0, 1, 2$,

$$L_{n,j} - \mathbb{E}[L_{n,j}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right)$$

and

$$Cov(L_{n,j}, L_{n,j'}) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right), \qquad j, j' = 0, 1, 2 \tag{3.13}$$

where $L_{n,j} = \dfrac{1}{nh^j\phi_x(h)} \displaystyle\sum_{i=1}^n \delta_i \beta^j(X_i, x) K(h^{-1} d(x, X_i))$.

Lemma 5. Based on the conditions of the Theorem 2, we get, for all $j = 0, 1, 2$,

$$Z_{n,j} - \mathbb{E}[Z_{n,j}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right)$$

and

$$Cov(Z_{n,j}, Z_{n,j'}) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right), \qquad j, j' = 0, 1, 2$$

where $Z_{n,j} = \dfrac{1}{n\ell h^j\phi_x(h)} \displaystyle\sum_{i=1}^n \delta_i \beta^j(X_i, x) K(h^{-1} d(x, X_i)) H(\ell_l^{-1}(y - Y_i))$.

Lemma 6. Under the same conditions that used in Theorem 2, we determine

$$\left|\widehat{B}(x)\right| = O\left(h^{\gamma_1}\right) + O\left(\ell^{\gamma_2}\right),$$

where $\widehat{B}(x) = \dfrac{\mathbb{E}[F_{12}(x,h)\ell_l^{-1}H(\ell_l^{-1}(y-Y_1))]}{\mathbb{E}[F_{12}(x,h)]} - \varphi(y|x)$

■

4. Computational aspects

4.1. Simulation result

We evaluate the efficiency of the delivered estimators, based on a finite sample, by comparing their behavior to the local constant, defined by

$$\widetilde{\varphi_n}(y|x) = \frac{\sum_{i=1}^{n} H\left(\ell_l^{-1}(y-Y_i)\right) K\left(\ell_l\|x-X_i\|\right)\delta_i}{\sum_{i=1}^{n} \ell_l K\left(h_k^{-1}\|x-X_i\|\right)\delta_i}.$$

For this empirical example, we draw a sample of random exploratory variable from the following continuous process

$$X(t) = \frac{Wt}{1+cos(Wt\pi)+W^2t^2}, \quad W \text{ is the standard normal distribution.}$$

Using the same grid, we discretize all the curves X_i's. Equi-spaced measurements of 100 in $(0,1)$ are used for generating the curves' grid. Shown in Figure 1 is a plot for all these functional curves.

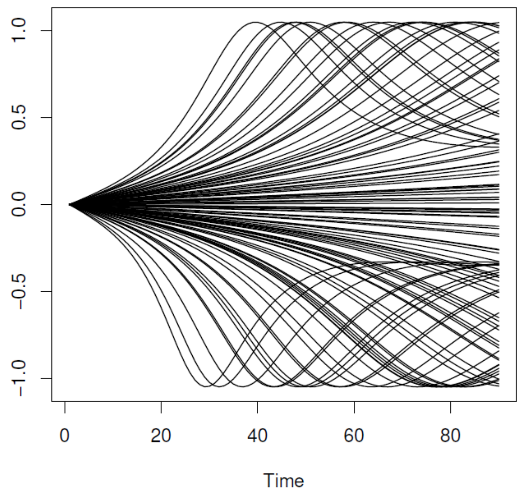

Figure 1. The results of plotting 100 samples curves.

Next, we generate the response variable by the relation,

$$Y = r(X) + \epsilon, \quad \text{with} \quad r(X) = \int_0^1 \frac{1}{1+X^2(t)}dt,$$

where ϵ is drown from the standard normal distribution. The CDP of $Y \mid X$ is clearly obtained by shifting the distribution of ϵ. It is expressed by

$$\varphi(y|x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-r(x))^2}{2}\right).$$

Of course, the novelty of this work is the examination of the behavior of the constructed estimators under various missing rates. Precisely, we wish to compare the resistance of φ_n, $\widehat{\varphi_n}$ and $\widetilde{\varphi_n}$ to the missing phenomena. For this purpose, we control this phenomena by the following conditional probability

$$P(X) = \ \texttt{expit}\left(\frac{2\alpha}{\pi}\int_0^1 X^2(t)dt\right),$$

where $\texttt{expit}(u) = e^u/(1+e^u))$. Such kind of conditional probability of observation has been used by [18] where the missing rate is measured by the parameter α. It is clear that the dependency to the regressor X increases with higher values of α which also decrease the degree of the missing rate. In this empirical study, we compare three missing rates: strong, medium and weak cases with $\alpha = 0.05$, $\alpha = 0.5$ and $\alpha = 5$ respectively. The missing rate is quantified by the following benchmark

$$\bar{\delta} = \frac{1}{n}\sum_{i=1}^{n}\delta_i.$$

The latter gives 5% missed observations for the weak case, 20% missed observations for the medium case and there are more than (55%) missing observations in the strong case. For the practical use of the estimators, we use the conditional mode cross-validation rule to select the different parameter involved in the estimators. This rule has been used by [19] for the prediction problem. It is based on the following criterion

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \bar{M_n}^{-i}(X_i)\right)^2, \quad \text{where} \quad \bar{M_n}^{-i}(X_i) = \arg\max_y \bar{\varphi_n}^{-i}(y|X_i), \tag{4.1}$$

with $\bar{\varphi_n}^{(-i)}$ refers to the estimation leave-one-out-curve of φ either φ_n, $\widehat{\varphi_n}$ or $\widetilde{\varphi_n}$. This rule is used to select the optimal number k and the threshold J of the Fourier basis. The first one is selected over $\{5, 10, 15, \ldots, 60\}^2$ and the second one from $\{2, 5, 8, 11, 12\}$. Finally, we used L_2 metric to simulate a quadratic kernel function that is supported in the interval $(0, 1)$. The efficiency of the three estimators is examined by the MAE-error expressed by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\varphi(Y_i|X_i) - \bar{\varphi_n}^{-i}(Y_i|X_i)\right|.$$

In Table 1, we summarize the MAE- error of the three estimators for various values of n. For more readability of the effect of the missing rate, we plot the true conditional densities versus their estimators for the three missing situations for $n = 100$ and for an arbitrary conditioning curves X_0 (randomly choosing); see Figure 2.

Table 1. MAE results.

		Weak case	Medium case	Strong case
MAE($\widehat{\varphi_n}$)	n=50	0.29	0.53	0.76
	n=150	0.23	0.41	0.67
	n=250	0.17	0.20	0.58
MAE(φ_n)	n=50	0.25	0.48	0.84
	n=150	0.22	0.37	0.66
	n=250	0.15	0.25	0.62
MAE($\widetilde{\varphi_n}$)	n=50	0.38	0.63	0.97
	n=150	0.32	0.43	0.69
	n=250	0.28	0.36	0.74

Clearly, the local linear approach appears to perform better than the classical kernel estimation of the CDP $\widetilde{\varphi_n}$, and the missing phenomena affects strongly the behavior of the estimation processing. Moreover, without surprise, the efficiency of the three estimators decreases with respect to the sample size n. We also observe that the three estimators have satisfactory results for $n = 50$ (see Table 1 which is a moderate sample size).

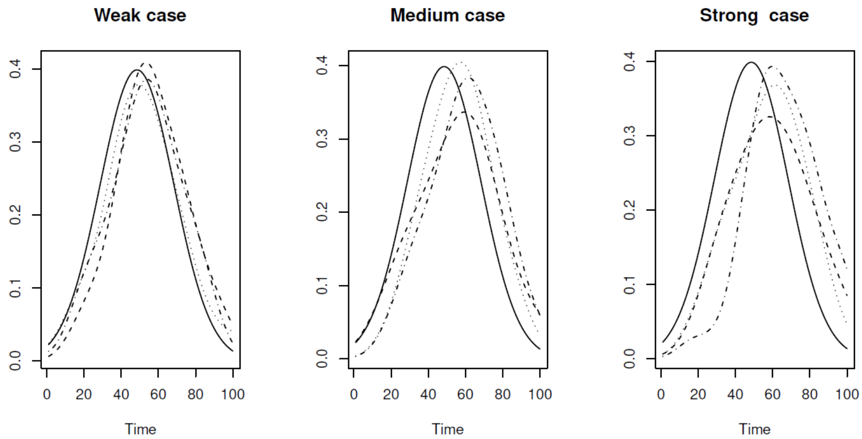

Figure 2. Comparison of the three estimators φ_n(—-), $\widehat{\varphi_n}$(.....) or $\widetilde{\varphi_n}$ (-.-.-).

4.2. A real-data application

In the real illustrative example, we examine how our approach performs over some food quality data. Specifically, we focus on the determination of riboflavin content in the yogurt using the Near-infrared curves. Such vitamin constitutes an important factor in yogurt quality. It is beneficial for the vision mechanism as well as the integrity of the skin and mucous membranes. The prediction of this vitamin has been considered by [29] using the multivariate statistical models. The novelty of the present works is the modelization by the nonparametric functional statistical models. They are more adapted for the functional nature of the spectrometry curves. In particular, in this real data analysis, we consider 115 curves $X_i(t)$ that represents the light absorption at 120 waves recorded from range excitation wavelengths ranged between 270 and 550 nm, and emission wavelengths ranged between 310-590. The data-set available at http://www.models.life.ku.dk/Yogurt and the curves are plotted in Figure 3.

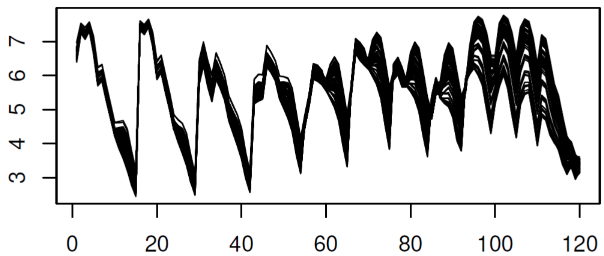

Figure 3. The spectra curves.

Of course the scalar response variable Y_i is the riboflavin content in each observed unite $i = 1, \dots 115$. Our main purpose is to examine the behavior of estimators φ_n and $\widehat{\varphi_n}$ in presence of the missing observations. To do that we compare the prediction by the conditional mode of the two estimators using two strategies to overcome the missing phenomena:

In the first strategy we omit the missing observation. In sense that, we predict an observation Y_{i_0} by

$$M_n(X_{i_0}) = \arg\max_y \varphi_n(y|X_{i_0}) \quad \text{and} \quad \widehat{M_n}(X_{i_0}) = \arg\max_y \widehat{\varphi_n}(y|X_{i_0}).$$

While, the second strategy is based on the estimation of the missing observation by the local mean. More precisely, we predict an observation Y_{i_0} by

$$M_n^{LM}(X_{i_0}) = \arg\max_y \varphi_n^{LM}(y|X_{i_0}) \quad \text{and} \quad \widehat{M_n}^{LM}(X_{i_0}) = \arg\max_y \widehat{\varphi_n}^{LM}(y|X_{i_0}),$$

where φ_n^{LM} and $\widehat{\varphi_n}^{LM}$ are computed by replacing the missing observation Y_{j_0} by

$$\bar{Y_{j_0}} = \frac{1}{k}\sum_{i=1}^{k} Y_{i,j_0},$$

$(Y_{i,j_0})_{i=1,\dots k}$ being the k response observations associated to the kNN explanatory observations at X_{j_0}. Furthermore, we keep the same parameters and the same rules of the simulation section. We change only the metric and the basis functions which are strongly affected by the smoothed property of the spectrometry curves. So, as the curves have discontinuous form, we proceeded with the PCA- projection basis and the PCA-metric where the best threshold of eigenfunction J is selected from the subset $\{1, 3, 5, 7, 9, 112\}$. We present the predication results in Table 2 and plotted them in Figure 4 and Figure 5. In these figures, the predicted values versus the real values are shown for 35 observations randomly chosen as testing points. This computational study is carried out over two levels of missing rates: weak (5% observations are omitted) and strong (50% observations are omitted).

Table 2. MSE results.

	Weak case	Strong case
MSE(M_n)	0.17	0.27
MSE($\widehat{M_n}$)	0.24	0.39
MSE(M_n^{LM})	0.48	0.67
MSE($\widehat{M_n}^{LM}$)	0.56	0.62

It appears clearly that the first strategy has more advantages over the second one. This conclusion confirms Efromovich's [15] statement that ignoring missing observations is the best strategy to handle the missing phenomena with nonparametric modeling. In addition, the two estimators are clearly very simple to apply in practice, and their efficiency is directly related to the selection of the different parameters included in the estimation methods.

5. Conclusion and prospects

In this paper, we have studied the problem of the nonparametric estimation of the conditional density function using the local linear approach. We have considered the kNN smoothing approach that allows us to improve the estimator's efficiency by selecting the appropriate bandwidth parameter. The second feature of this study is the possibility to cover the incomplete data situation characterized by the missing phenomena. Empirical analysis shows the excellent performance of the proposed methodology, which varied with respect to the missing level. In addition to these features, the present study opens some crucial tracks for the future. In particular, it will be interesting to investigate the other types of incomplete functional data, such as the censored or truncated data. Another possible direction is to study the asymptotic property of the kNN local linear estimator in the functional times series case (complete or incomplete cases). In addition, the asymptotic distribution of the proposed estimators is an interesting open question. Such asymptotic property is essential as preliminary statistical analyses, including the confidence interval or hypotheses testing. In addition, extending this kind of estimation to other nonparametric models, such as the conditional hazard function or the conditional distribution function, is also a natural prospect of the present contribution.

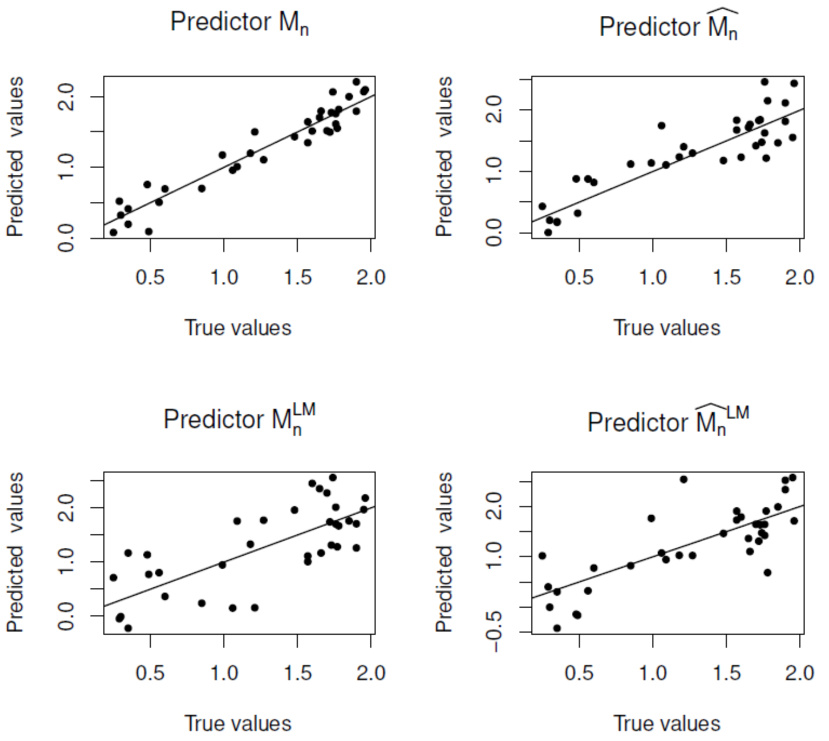

Figure 4. Weak case.

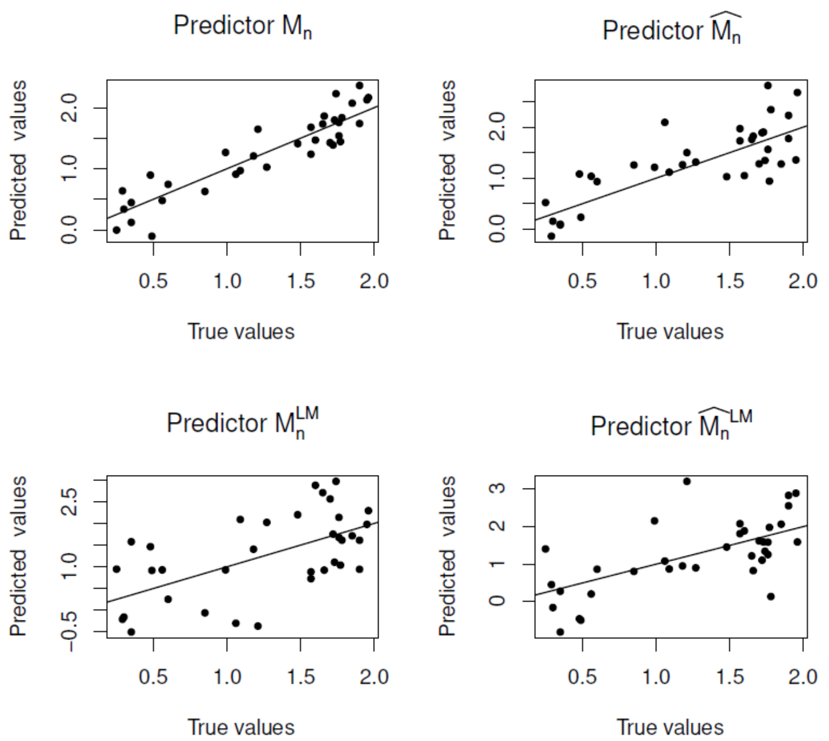

Figure 5. Strong case.

Acknowledgment. The authors would like to thank the Editor, an Associate-Editor and an anonymous reviewer, for their valuable comments and suggestions which improved substantially the quality of an earlier version of this paper. The authors also thank and extend their appreciation to the Deanship of Scientific Research at King Khalid University, Kingdom of Saudi Arabia for supporting and funding this work through the research groups program under the project number R.G.P.1/177/43.

References

[1] I.M. Almanjahie, Z. Chikr Elmezouar, B.A. Bachir, and Z. Kaid, *Spatial local linear estimation of the L-1-conditional quantiles for functional regressors*, Comm. Statist. Theory Methods **49** (23), 5666-5685, 2020.

[2] I.M. Almanjahie, Z. Chikr Elmezouar, A. Laksaci and M. Rachdi, *kNN local linear estimation of the conditional cumulative distribution function: Dependent functional data case*, C. R. Math. **356** (10), 1036-1039, 2018.

[3] G. Aneiros Pérez, R. Cao and P. Vieu, *Editorial on the special issue on functional data analysis and related topics*, Comput. Statist. **34** (2), 447-450, 2019.

[4] M. Attouch and F. Belabed, (2014), *The k nearest neighbors estimation of the conditional hazard function for functional data*, REVSTAT **12** (3), 273-297, 2014.

[5] M. Attouch and W. Bouabça, *The k-nearest neighbors estimation of the conditional mode for functional data*, Roumaine Math. Pures Appl. **58** (4), 393-415, 2013.

[6] A. Baìllo and A. Grané, *Local linear regression for functional predictor and scalar response*, J. Multivariate Anal. **100** (1), 102-111, 2009.

[7] J. Barrientos-Marin, F. Ferraty and P. Vieu, *Locally modelled regression and functional data*, J. Nonparametr. Stat. **22** (5), 617-632, 2010.

[8] A. Benchiha and Z. Kaid, *Local linear estimate for functional regression with missing data at random*, Int. J. Math. Stat. **19**, 22-33, 2018.

[9] E. Boj, P. Delicado and J. Fortiana, *Distance-based local linear regression for functional predictors*, Comput. Statist. Data Anal. **54** (2), 429-437, 2010.

[10] F. Burbea, F. Ferraty and P. Vieu, *k-nearest neighbor method in functional nonparametric regression*, J. Nonparametr. Stat. **21** (4), 453-469, 2009.

[11] Z. Chikr Elmezouar, I.M. Almanjahie, A. Laksaci and M. Rachdi, *FDA: strong consistency of the kNN local linear estimation of the functional conditional density and mode*, J. Nonparametr. Stat. **31** (1), 175-195, 2019.

[12] G. Collomb, W. Härdle and S. Hassani, *A note on prediction via estimation of the conditional mode function*, J. Statist. Plann. Inference **15**, 227-236, 1987.

[13] S. Dabo-Niang, Z. Kaid and A. Laksaci, *Asymptotic properties of the kernel estimate of spatial conditional mode when the regressor is functional*, AStA Adv. Stat. Anal. **99** (2), 131-160, 2015.

[14] J. Demongeot, A. Laksaci, F. Madani and M. Rachdi, *Functional data: local linear estimation of the conditional density and its application*, Statistics **47** (1), 26-44, 2013.

[15] S. Efromovich, *Missing and modified data in nonparametric estimation with R examples*, in Monographs on Statistics and Applied Probability, 156, CRC Press, 2018.

[16] M. Ezzahrioui and E. Ould Saïd, *Some asymptotic results of a non-parametric conditional mode estimator for functional time-series data*, Stat. Neerl. **64** (2), 171-201, 2010.

[17] F. Ferraty, A. Laksaci and P. Vieu, *Estimating some characteristics of the conditional distribution in nonparametric functional models*, Stat. Inference Stoch. Process. **9** (1), 47-76, 2006

[18] F. Ferraty, M. Sued and P. Vieu, *Mean estimation with data missing at random for functional covariables*, Statistics **47** (4), 688-706, 2013.

[19] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer-Verlag, 2006.

[20] L. Kara-Zaitri, A. Laksaci, M. Rachdi and P. Vieu, *Data-driven kNN estimation for various problems involving functional data*, J. Multivariate Anal. **153**, 176-188, 2017.

[21] N. Kudraszow, and P. Vieu, *Uniform consistency of kNN regressors for functional variables*, Statist. Probab. Lett. **83** (8), 1863-1870, 2013.

[22] A. Laksaci, *Quadratic error of the kernel estimator of conditional density when the regressor is functional*, C. R. Math. Acad. Sci. Paris **345** (3), 171-175, 2007.

[23] A. Laksaci and A. Yousfate, *Functional estimate of Markov transition operator density: discrete time case*, C. R. Math. Acad. Sci. Paris **334** (11), 1035-1038, 2002.

[24] H. Lian, *Convergence of functional k-nearest neighbor regression estimate with functional responses*, Electron. J. Stat. **5**, 31-40, 2011.

[25] N. Ling, Y. Liu and P. Vieu, *Nonparametric regression estimation for functional stationary ergodic data with missing at random*, J. Statist. Plann. Inference **162**, 75-87, 2015.

[26] N. Ling, Y. Liu and P. Vieu, *Conditional mode estimation for functional stationary ergodic data with responses missing at random*, Statistics **50** (5), 991-1013, 2016.

[27] N. Ling, and P. Vieu, *Nonparametric modelling for functional data: selected survey and tracks for future*, Statistics **52** (4), 934-949, 2018.

[28] D. Louani, and E. Ould-Saïd, *Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis*, J. Nonparametr. Stat. **11** (4), 413-442, 1999.

[29] E. Miquel Becker, J. Christensen, C.S. Frederiksen and V.K Haugaard, *Front-face fluorescence spectroscopy and chemometrics in analysis of yogurt: rapid analysis of riboflavin*, J. Dairy Sci. **86** (8), 2508-2515, 2003.

[30] A. Quintela-Del-Río and P. Vieu, *A nonparametric conditional mode estimate*, J. Nonparametr. Stat. **8** (3), 253-266, 1997.

[31] M. Rachdi, A. Laksaci, I.M Almanjahie, and Z. Chikr Elmezouar, *FDA: theoretical and practical efficiency of the local linear estimation based on the kNN smoothing of the conditional distribution when there are missing data*, J. Stat. Comput. Simul. **90** (8), 1479-1495, 2020.

[32] M. Rachdi, A. Laksaci, J. Demongeot, A. Abdali and F. Madani, *Theoretical and practical aspects of the quadratic error in the local linear estimation of the conditional density for functional data*, Comput. Statist. Data Anal. **73**, 53-68, 2014.

[33] G.G. Roussas, *Nonparametric estimation in Markov processes*, Ann. Inst. Statist. Math. **21** (1), 73-87, 1969.

[34] J.V. Uspensky, (1937). *Introduction to Mathematical Probability*, McGraw-Hill Book Company, Inc., 1937.

[35] É. Youndjé, P. Sarda and P. Vieu, *Validation croisée pour lestimation non-paramétrique de la densité conditionnelle*, Publ. Inst. Statist. Univ. Paris **38** (1), 57-80, 1994.

Appendix

The proofs of the intermediate results are given in short ways because we follow the same ideas as in [11]. The main challenge, here, is how to handle the additional variable δ. When no ambiguity is necessary then in what follows, we will denote some strictly positive generic constants by C and C^*.

Proof of Lemma 2. For the first term, we write

$$\mathbb{E}[T_{n,j}] = \frac{1}{\ell\phi_x(h)}\mathbb{E}\left[c_{1j}\delta_1 K(h^{-1}\|x-X_i\|)\left(H(\ell^{-1}(y-Y_i)) - \varphi(y|X_1)\right)\right].$$

Conditioning by X_1, to conclude that

$$\mathbb{E}[T_{n,j}] = \frac{1}{h\ell\phi_x(h)}\mathbb{E}\left[c_{1j}K(h^{-1}\|x-X_1\|)P(X_1)(\mathbb{E}\left[H(\ell_l^{-1}(y-Y_i))|X_1\right] - \varphi(y|X_1))\right].$$

Observe that by standard analytical arguments, we obtain

$$\mathbb{E}\left[\ell^{-1}H(\ell_l^{-1}(y-Y_i))|X_1\right] = \varphi(y|X_1)) + O(\ell^{\gamma}).$$

It follows that

$$\mathbb{E}[T_{n,j}] = O(\ell^{\gamma}).$$

The proof of Lemma 2 is now completed, because the second term is treated in the same manner.

Proof of Lemma 3. The main tool for proofing this lemma is the use of Bernstein's inequality on

$$\Delta_i = \frac{1}{h\ell\phi_x(h)}\delta_i c_{ij}K(h^{-1}\|x-X_i\|)P(X_1)(\mathbb{E}\left[H(\ell^{-1}(y-Y_i))|X_1\right] - \varphi(y|X_1)),$$

for which

$$T_{n,j} = \frac{1}{n}\sum_{i=1}^{n}[\Delta_i - \mathbb{E}[\Delta_i]].$$

Now, because of $(v_j)_{j\geq 1}$ is an orthonormal basis, for all $j \leq J$, we obtain

$$|c_{1j}| \leq \|v_j\|\|x-X_1\| \leq \|x-X_1\|.$$

Thus,

$$\mathbb{E}\left[c_{ij}\delta_1 K_1(h)\right] \leq \mathbb{E}\left[\|x-X_1\|^2 K_1(h)\right] \leq Ch\phi_x(h).$$

By using conditions (2.2) and (3.3), we obtain that

$$|\Delta| < C/\ell\phi_x(h) \quad \text{and} \quad \mathbb{E}\,|\Delta_i|^2 < C'/\ell\phi_x(h).$$

Then, the Bernstein inequality (see [34], Page 205) permits to infer, for $\eta > 0$, that

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(\Delta_i - \mathbb{E}[\Delta_i])\right| > \eta\sqrt{\frac{\log n}{n\,\ell\phi_x(h)}}\right\} \;\leq\; C^* n^{-C\eta^2}.$$

We conclude that

$$T_{n,j} - \mathbb{E}[T_{n,j}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\,\ell\phi_x(h)}}\right).$$

The proof of $e_{n,j}$ follows the same line as $T_{n,j}$.

Proof of lemma 4. Because the variable δ and the kernel K are bounded, we obtain, for all $i = 1, \dots n$

$$\left|\frac{1}{h^j\phi_x(h)}\delta_i\beta^j(X_i,x)K(h^{-1}d(x,X_i))\right| \leq C\frac{1}{\phi_x(h)}$$

$$\text{and} \quad Var\left[\frac{1}{h^j\phi_x(h)}\delta_i\beta^j(X_i,x)K(h^{-1}d(x,X_i))\right] \leq C'\frac{1}{\phi_x(h)},$$

which allows to infer that

$$\mathbb{P}\left\{|L_{n,j} - \mathbb{E}[L_{n,j}]| > \eta\sqrt{\frac{\log n}{n\,\phi_x(h)}}\right\} \;\leq\; C^* n^{-C\xi^2}, \quad \text{for certain} \quad \xi > 0.$$

On the other hand, by the same reasoning we obtain for all $l = 0, 1, 2$, and $k = 0, 1$, we have

$$\mathbb{E}\left[\delta_i\beta^l(X_i,x)K^k(h^{-1}d(x,X_i))\right] = O(h^l\phi_x(h)). \tag{A.1}$$

This last assertion is a consequence of the missing property where

$$\mathbb{E}\left[\delta_i\beta^j(X_i,x)K^k(h^{-1}d(x,X_i))\right] = \mathbb{E}\left[P(X_i)\beta^j(X_i,x)K^k(h^{-1}d(x,X_i))\right].$$

Now, concerning the covariance term, we use the independence and the stationarity of the $(X_i, \delta_i, Y_i)_i$ to write that

$$\begin{aligned} Cov(L_{n,j}, L_{n,j'}) &= \frac{1}{nh^{j+j'}\phi_x^2(h)}Cov(\delta_1\beta^j(X_1,x)K(h^{-1}d(x,X_1)), \delta_1\beta^{j'}(X_1,x)K(h^{-1}d(x,X_1))) \\ &= \frac{1}{nh^{j+j'}\phi_x^2(h)}\mathbb{E}\left[\delta_1^2\beta^{j+j'}(X_i,x)K^2(h^{-1}d(x,X_i))\right] \\ &- \frac{1}{nh^{j+j'}\phi_x^2(h)}\mathbb{E}\left[\delta_1\beta^j(X_i,x)K(h^{-1}d(x,X_i))\right]\mathbb{E}\left[\delta_1\beta^{j'}(X_i,x)K(h^{-1}d(x,X_i))\right] \end{aligned}$$

Finally, using to deliver the proof of Equation (3.13).

Proof of Lemma 5. Proofing this lemma is very comparable to that of Lemma 4. The key difference is in the additional term $\ell^{-1}H^j(\ell^{-1}(y - Y_i))$; for $j = 1, 2$. The latter can be manipulated by standard arguments. Indeed, we write

$$\mathbb{E}\left[\ell^{-1}H^j(\ell^{-1}(y - Y_i))|X\right] = \int_{\mathbb{R}} H^j(t)\varphi(y - \ell t|X)dt.$$

Because of the Holderian condition (2.4) and the condition (3.5) on the kernel H, we prove that

$$\mathbb{E}\left[\ell^{-1}H^j(\ell^{-1}(y - Y_i))|X\right] = \int_{\mathbb{R}} H^j(t)\varphi(y - \ell t|X)dt = O(1).$$

Hence, for $l = 0, 1, 2$, $j = 1, 2$ and $k = 0, 1, 2$, we determine that

$$\mathbb{E}\left[\delta_i\beta^l(X_i,x)K^k(h^{-1}d(x,X_i))\ell^{-j}H^j(\ell_l^{-1}(y - Y_i))\right] = O(h^l\ell^{j-1}\phi_x(h)).$$

Thus, to achieve the first part, we employ Bernstein's inequality to conclude that

$$Z_{n,j} - \mathbb{E}[Z_{n,j}] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\,\phi_x(h)}}\right).$$

Note that the stationarity property for the observations $(X_i, \delta_i, Y_i)_i$ can be used to conclude that

$$Cov(Z_{n,j}, Z_{n,j'}) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right) \qquad j, j' = 0, 1, 2.$$

Proof of Lemma 6. We start with

$$\begin{aligned}\widehat{B}(x) &= \frac{\mathbb{E}[F_{12}(x,h)\ell^{-1}H(\ell^{-1}(y-Y_2))]}{\mathbb{E}[F_{12}(x,h)]} - \varphi(y|x) \\ &= \frac{1}{\mathbb{E}[F_{12}(x,h)]}\mathbb{E}\left[F_{12}(x,h)(\ell^{-1}H(\ell^{-1}(y-Y_2)) - \varphi(y|x))\right].\end{aligned}$$

Then, treating the nominator term as

$$\begin{aligned}&\mathbb{E}\left[F_{12}(x,h)(\ell^{-1}H(\ell^{-1}(y-Y_2)) - \varphi(y|x))\right] \\ &\quad = \mathbb{E}\Big[\delta_1\beta^2(X_1,x)K(h^{-1}d(x,X_1))K(h^{-1}d(x,X_2))P(X_2) \\ &\qquad (\mathbb{E}[\ell^{-1}H(\ell^{-1}(y-Y_2))] - \varphi(y|x)) \\ &\qquad -\delta_1\beta(X_1,x)K(h^{-1}d(x,X_1))\beta(X_2,x)K(h^{-1}d(x,X_2))P(X_2) \\ &\qquad (\mathbb{E}[\ell^{-1}H(\ell^{-1}(y-Y_2))] - \varphi(y|x))\Big].\end{aligned}$$

Once again, we use the Holderian condition (2.4) to show that

$$\mathbb{E}\left[F_{12}(x,h)(\ell^{-1}H(\ell^{-1}(y-Y_2)) - \varphi(y|x))\right] = O\left(h^{\gamma_1}\right) + O\left(\ell^{\gamma_2}\right).$$

Therefore,

$$\widehat{B}(x) = O\left(h^{\gamma_1}\right) + O\left(\ell^{\gamma_2}\right).$$

Proof of Theorems For sake of brevity we proof only how the convergence rate of Theorem 1 results from the Lemmas 1-3. The proof of Theorem 2 follows the same lines. Indeed, as the convergence rates of Lemmas 1-3 are stated for h equal h_k^r and h_k^l (respectively ℓ by ℓ^r and ℓ^l), then , if we replace in Lemma2 h by

$$h_k^r = \phi_x^{-1}\left(k/\alpha_1 n\right) \quad \left(\text{resp.} h_k^l = \phi_x^{-1}\left(\alpha_2 k/n\right)\right)$$

and ℓ by

$$\ell^r = \frac{\alpha_3 l}{n} \qquad \text{and} \qquad \ell^l = \frac{l}{\alpha_3 n}$$

we get

$$\mathbb{E}\left[T_{n,j}\right] = O\left(\frac{\alpha_3^\gamma l^\gamma}{n^\gamma}\right) = O\left(\frac{l}{n}\right)^\gamma \text{ for } \ell = \ell^r$$

and

$$\mathbb{E}\left[T_{n,j}\right] = O\left(\frac{l^\gamma}{\alpha_3^\gamma n^\gamma}\right) = O\left(\frac{l}{n}\right)^\gamma \text{ for } \ell = \ell^l.$$

Thus, for both cases ℓ by ℓ^r and ℓ^l), we have

$$\mathbb{E}\left[T_{n,j}\right] = O\left(\frac{l}{n}\right)^\gamma$$

which is exactly the second term in Theorem 1. Using the same reasoning for $e_{n,j}$. Indeed, as α_1, α_2 such that

$$\phi_x^{-1}\left(\frac{k}{\alpha_1 n}\right) \le C\phi_x^{-1}\left(\frac{\alpha_2 k}{n}\right),$$

hence,

$$\mathbb{E}\left[e_{n,j}\right] = O\left(\left(\phi_x^{-1}\left(k/\alpha_1 n\right)\right)^2\right) = O\left(\left(\phi_x^{-1}\left(\alpha_2 k/n\right)\right)^2\right) \text{for } \ h = h^r$$

and

$$\mathbb{E}\left[e_{n,j}\right] = O\left(\left(\phi_x^{-1}\left(\alpha_2 k/n\right)\right)^2\right) \text{for } \ h = h^l$$

On the other hand, because of ϕ is increasing function we have, for $\alpha_2 \in (0,1)$,

$$\phi_x^{-1}\left(\alpha_2 k/n\right) \leq \phi_x^{-1}\left(k/n\right).$$

We conclude that

$$\mathbb{E}\left[e_{n,j}\right] = O\left(\left(\phi_x^{-1}\left(k/n\right)\right)^2\right) \text{for } \ h = h^r$$

and

$$\mathbb{E}\left[e_{n,j}\right] = O\left(\left(\phi_x^{-1}\left(k/n\right)\right)^2\right) \text{for } \ h = h^l$$

which is exactly the third term in Theorem 1. Now, for the last term in Theorem 1, we use the same ideas as in Lemma 2. Indeed, firstly observe that

$$\phi_x(h) = \phi_x\left(\phi_x^{-1}\left(\alpha_1 k/n\right)\right) = \frac{\alpha_1 k}{n} \text{for } \ h = h^r$$

and

$$\phi_x(h) = \phi_x\left(\phi_x^{-1}\left(k/\alpha_2 n\right)\right) = \frac{k}{\alpha_2 n} \text{for } \ h = h^l.$$

Then, the convergence rates in Lemma 3 are

$$T_{n,j} - \mathbb{E}\left[T_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\frac{\alpha_3 l}{n}\frac{\alpha_1 k}{n}}}\right) = O_{a.co.}\left(\sqrt{\frac{n\log n}{lk}}\right)$$
$$\text{for } \ h = h^r,\ \ell = \ell^r$$

and

$$T_{n,j} - \mathbb{E}\left[T_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\ell\phi_x(h)}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\frac{l}{\alpha_3 n}\frac{k}{\alpha_2 n}}}\right) = O_{a.co.}\left(\sqrt{\frac{n\log n}{lk}}\right)$$
$$\text{for } \ h = h^l,\ \ell = \ell^l.$$

Similarly,

$$e_{n,j} - \mathbb{E}\left[e_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\frac{\alpha_1 k}{n}}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right) \text{for } \ h = h^r$$

and

$$e_{n,j} - \mathbb{E}\left[e_{n,j}\right] = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(h)}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{n\frac{k}{\alpha_2 n}}}\right) = O_{a.co.}\left(\sqrt{\frac{\log n}{k}}\right) \text{for } \ h = h^l.$$

It is clear that leading term between two convergence rate of $T_{n,j}$ and $e_{n,j}$ is $O_{a.co.}\left(\sqrt{\frac{n\log n}{lk}}\right)$ which is exactly the last term in Theorem 1.