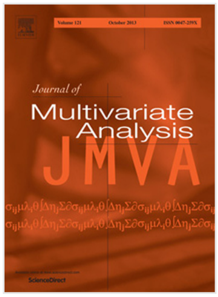
Accepted Manuscript



Data-driven kNN estimation in nonparametric functional data analysis

Lydia-Zaitri Kara, Ali Laksaci, Mustapha Rachdi, Philippe Vieu

PII: S0047-259X(16)30110-5
DOI: http://dx.doi.org/10.1016/j.jmva.2016.09.016
Reference: YJMVA 4169

To appear in: *Journal of Multivariate Analysis*

Received date: 20 February 2016

Please cite this article as: L.-Z. Kara, A. Laksaci, M. Rachdi, P. Vieu, Data-driven kNN estimation in nonparametric functional data analysis, *Journal of Multivariate Analysis* (2016), http://dx.doi.org/10.1016/j.jmva.2016.09.016

Data-driven kNN estimation in nonparametric functional data analysis

Lydia-Zaitri Kara[a], Ali Laksaci[a], Mustapha Rachdi[1b], Philippe Vieu[c]

[a] *Université Djillali Liabès, LSPS, Sidi Bel-Abbès, Algeria, karazaitri@hotmail.fr and alilak@yahoo.fr*
[b] *Université Grenoble Alpes, AGIM Team, AGEIS EA 7407, France, mustapha.rachdi@univ-grenoble-alpes.fr*
[c] *Université Paul-Sabatier, IMT, Toulouse, France, vieu@math.univ-toulouse.fr*

Abstract

Kernel nearest-neighbor (kNN) estimators are introduced for the nonparametric analysis of statistical samples involving functional data. Asymptotic theory is provided for several different target operators including regression, conditional density, conditional distribution and hazard operators. The main point of the paper is to consider data-driven methods of selecting the number of neighbors in order to make the proposed methods fully automatic. As a by-product of our proofs we state consistency results for kNN functional estimators which are uniform in the number of neighbors (UINN). Some simulated experiences illustrate the feasibility and the finite-sample behavior of the method.

Keywords: Functional data analysis; UINN consistency; Functional nonparametric statistics; kNN estimator; Data-driven estimator.

Subject classifications: 62G05, 62G08, 62G20, 62G35, 62G07, 62G32, 62G30. Secondary: 62H12

1. Introduction

In view of their flexibility and efficiency, nonparametric k-nearest-neighbor (kNN) smoothing methods have received a great attention in the statistical

[1] Corresponding author: M. Rachdi, Université Grenoble Alpes, UFR SHS, BP. 47, 38040 Grenoble Cedex 09, France.

literature for analyzing multivariate data. Work in this direction was initiated by Cover [12] and a very large set of papers is now available in various estimation contexts such as regression [11, 16, 40] discrimination [18, 31], density estimation [3, 17, 35] and clustering analysis [41]. The book by Györfi et al. [25] provides an extensive study of kNN estimators in the finite dimensional setting.

One of the main interest of the kNN approach, compared with classical kernel estimators, is the fact that it includes a locally adaptive smoothing parameter allowing for the control of local heterogeneity in the data. Because the local structures in the data are more and more influent when the dimension increases (see, e.g., [39]) the kNN approach is particularly well adapted to multivariate problems. In infinite-dimensional problems the need for constructing location adaptive estimators is even more crucial; see [4] for empirical studies. Therefore the kNN ideas are expected to lead to attractive statistical methods for functional data analysis, and this is the point we want to address in this work.

Recently, statistical inference for Functional Data Analysis (FDA) has been deeply investigated; see [6, 26, 27, 38, 44] for selected general books on this topic. Nonparametric ideas have been popularized by the book of Ferraty and Vieu [21] and now take a large place in the FDA literature; see the specific discussions in the recent surveys by Cuevas [13] and Goia and Vieu [24]. In particular, there is an extensive list of contributions in nonparametric functional statistics concerning standard kernel estimators; see, e.g., [9, 20, 22, 34, 36] for a selected set of works in this direction. However, the study of kNN methods is still rather limited and mainly oriented towards the regression estimation [5, 7, 29, 30, 32] or towards curves discrimination [8]. Moreover, it should be stressed that the scant literature on functional kNN methods concerns a fixed number of neighbors while this number needs to be data-driven in practice.

The main purpose of our paper is then to state a wide scope of asymptotic results, covering several different target operators (regression, conditional density, conditional distribution function, conditional hazard function) and allowing for an automatic choice data-driven of the number of neighbors. The main tool for achieving this goal is the statement of results that are uniform in the number of neighbors (UINN). Thus, we establish the UINN almost-complete convergence (a.co.[2]). In fact, the UINN feature allows to control the asymptotic behavior of the estimators even if the number of neighbors is random, leading to direct applications to data-driven selected numbers of neighbors. As far as we know, this contribution is the first one to address the UINN properties for kNN estimators since, even in the standard multivariate case, this kind of literature is only concerned with kernel estimators. We refer to [15] for a list of references on the uniform in bandwidths (UIBB) consistency of traditional multivariate kernel estimators and to [28] for the functional extension.

This paper is organized as follows. We present our models and their estimators in Section 2. As a preliminary result, we give in Section 3 some asymptotics for a wide scope of kNN estimators leading directly to the main results of our paper (see Section 4) about the asymptotic behavior of the kNN estimators using a data-driven random number of neighbors. The technical proofs are postponed to Section 6. Some simulated examples are presented in Section 5 in order to stress the easy implementation of the method and its nice predictive behavior in finite samples. Finally, Section 7 is devoted to comments on the results and to related tracks for future.

2. Models and Estimators

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of independent and identically distributed pairs as (X, Y) which is a random vector valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space. In the following, d is a semi-metric on $\mathcal{F}$, x is a fixed point in $\mathcal{F}$, N_x is a fixed neighborhood of x, and the closed ball centered at x and of

[2] Let $z_1, z_2, \ldots$ be a sequence of real random variables. We say that (z_n) converges almost-completely (a.co.) to 0 if, and only if, for all $\epsilon > 0$, $\sum_{n=1}^{\infty} \Pr(|z_n| > \epsilon) < \infty$. Moreover, we say that the rate of the almost-complete convergence of (z_n) to zero is of order u_n (with $n \to 0$) and we write $z_n = O_{a.co.}(u_n)$ if, and only if, there exists $\epsilon > 0$ such that $\sum_{n=1}^{\infty} \Pr(|z_n| > \epsilon u_n) < \infty$. This kind of convergence implies both almost-sure convergence and convergence in probability.

radius α is denoted

$$B(x,\alpha) = \{y \in \mathcal{F} \text{ such that } d(y,x) \leq \alpha\}\,.$$

We study the asymptotic properties of the kNN kernel estimators in three nonparametric conditional models. The first one is the regression model

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

This model has been widely studied in the functional context; see [22] for recent advances. In most of the existing works, the relationship between the nonparametric model and the data space is translated through the following Lipschitz condition. Assume that for some constants $\beta > 0$ and $C_1 > 0$ we have

$$\forall_{x_1,x_2\in N_x} \quad |m(x_1) - m(x_2)| \leq C_1 d^{\beta_1}(x_1,x_2). \tag{2}$$

The functional kNN regression estimator is defined by

$$\widehat{m}(x) = \frac{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x,X_i)\} Y_i}{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x,X_i)\}}$$

where

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

and $\mathbf{1}_A$ is the indicator function of the set A.

The second statistical operator that we intend to study is the conditional probability distribution function which is defined as follows:

$$F^x(\cdot) = \Pr(Y \leq \cdot|X = x). \tag{3}$$

In fact F^x can be viewed as a particular case of the regression function m, where the response Y is replaced by the indicator function $\mathbf{1}_{\{Y\leq\cdot\}}$. Thus, the functional kNN estimator of F^x can be defined as

$$\widehat{F}^x(y) = \frac{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x,X_i)\}\mathbf{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x,X_i)\}} \tag{4}$$

and the nonparametric model is characterized by assuming that for some $\beta_2 > 0$ and some $C_2 > 0$ one has

$$\forall_{x_1,x_2\in N_x}\ \forall_{y\in\mathbb{R}} \quad |F^{x_1}(y) - F^{x_2}(y)| \leq C_2 d(x_1,x_2)^{\beta_2}. \tag{5}$$

Similarly, we also consider the conditional density function defined, at each point where F^x is differentiable, by

$$f^x(\cdot) = (F^x)'(\cdot) \tag{6}$$

and for which the kNN estimator[3] is defined by

$$\widehat{f^x}(y) = L_{\ell,y}^{-1} \frac{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x, X_i)\} K(L_{\ell,y}^{-1}|y - Y_i|)}{\sum_{i=1}^n K\{H_{k,x}^{-1} d(x, X_i)\}}$$

where

$$L_{\ell,y} = \min \left\{ z \in \mathbb{R}^+ \text{ such that } \sum_{i=1}^n \mathbf{1}_{(y-z,y+z)}(Y_i) = \ell \right\}$$

and where ℓ is a sequence of integers belonging to $(\ell_{1,n}, \ell_{2,n})$. The nonparametric aspect of this model is characterized by assuming that, for all $x_1, x_2 \in N_x$ and for all y_1, y_2 in a fixed neighborhood of y we have, for some $C_3, \beta_3, \beta_4 > 0$, that

$$|f^{x_1}(y_1) - f^{x_2}(y_2)| \leq C_3 \left\{ d(x_1, x_2)^{\beta_3} + |y_1 - y_2|^{\beta_4} \right\}. \tag{7}$$

It is obvious that, by combining the two last estimators, the kNN method can be used for the estimation of the hazard operator, viz.

$$r^x(\cdot) = \frac{f^x(\cdot)}{1 - F^x(\cdot)} \tag{8}$$

and the corresponding estimator is defined for all $y \in \mathbb{R}$ such that $F^x(y) < 1$, by

$$\widehat{r^x}(y) = \frac{\widehat{f^x}(y)}{1 - \widehat{F^x}(y)}.$$

3. UINN asymptotics

We start by gathering together all assumptions required to obtain our asymptotic results.

[3] To avoid tedious additional notations, we use the same kernel function K for weighting the functional variable X and the scalar one Y, but the results stated later on in this paper are also true for two different kernels.

(H1) For all $r > 0$, $\Pr\{X \in B(x, r)\} =: \phi_x(r) > 0$ such that, for all $s \in (0, 1)$,

$$\lim_{r \to 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s).$$

(H2) The class of functions

$$\mathcal{K} = \{\cdot \mapsto K\{\gamma^{-1} d(x, \cdot)\},\ \gamma > 0\} \text{ is a point-wise measurable class}^4$$

such that

$$\sup_Q \int_0^1 \sqrt{1 + \ln \mathcal{N}(\epsilon \|F\|_{Q,2}, \mathcal{K}, d_Q)} d\epsilon < \infty,$$

where the supremum is taken over all probability measures Q on the space $\mathcal{F}$ with $Q(F^2) < \infty$ and where F is the envelope function[5] of the set $\mathcal{K}$. Here, d_Q is the $L_2(Q)$-metric and $\mathcal{N}(\epsilon, \mathcal{K}, d_Q)$ is the minimal number of open balls (with respect to the $L_2(Q)$-metric) with radius ϵ which are needed to cover the function class $\mathcal{K}$. We will denote by $\|\cdot\|_{Q,2}$ the $L_2(Q)$-norm.

(H3) The kernel K is supported within $(0, 1/2)$ and has a continuous first derivative on $(0, 1/2)$ which is such that

$$0 < C_4 \mathbf{1}_{(0,1/2)}(\cdot) \leq K(\cdot) \leq C_5 \mathbf{1}_{(0,1/2)}(\cdot)$$

and

$$K(1/2) - \int_0^{1/2} K'(s) \tau_x(s) ds > 0,$$

(H4) The sequence of numbers $(k_{1,n})$ satisfies

$$\frac{\ln n}{\min\left\{n \phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n}\right\}} \to 0.$$

[4] A class of functions $\mathcal{C}$ is said to be a point-wise measurable class if there exists a countable subclass $\mathcal{C}_0$ such that for any function $g \in \mathcal{C}$ there exists a sequence of functions $(g_n)_{n \in \mathbb{N}}$ in $\mathcal{C}_0$ such that: $|g_n(z) - g(z)| = o(1)$.

[5] An envelope function G for a class of functions $\mathcal{C}$ is any measurable function such that $\sup_{g \in \mathcal{C}} |g(z)| \leq G(z)$ for all z.

We mention that most of our assumptions are standard in the functional nonparametric context; see, Ferraty and Vieu [21]. The only condition which is specific to our UINN purpose is Assumption (H2). The first part of (H2) is a mild measurability restriction which will allow to state uniform results. Observe that an alternative way to deal with this problem would be to deal with the notions of outer probability (see [42], p. 4) but this would need much more complicated computations. Moreover, for obtaining the UINN consistency, it is worth noticing that uniform asymptotics are closely related with the notions of entropy and compactness. The second part in condition (H2) is a uniform integral entropy condition used to characterize the Donsker-class of functions [42] and allows to derive a uniform limit distribution. Note that this kind of assumption could also be useful for evaluating moments of empirical processes [42]. The great generality of Assumption (H2) is further underlined by the fact that it is less restrictive than the usual VC-class condition [19].

Theorem 3.1. *Under Assumptions (H1)–(H4) we have*

$$\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_1}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right), \quad (9)$$

and if $\mathrm{E}(|Y|^q|X) < C_5 < \infty$*, almost-surely for some* $q \ge 2$ *and* $C_5 > 0$*, then we have*

$$\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{m}(x) - m(x)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right).$$

Furthermore, if we replace Assumption (H4) by the following (H4')

$$\frac{n\ln n}{\ell_{1,n}\min\left\{n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n}\right\}} = o(1),$$

then we obtain

$$\sup_{k_{1,n}\le k\le k_{2,n}} \sup_{\ell_{1,n}\le \ell\le \ell_{2,n}} |\widehat{f}^x(y) - f^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_2}\right\} + O\left(\frac{\ell_{2,n}}{n}\right)^{\beta_3} + O_{a.co.}\left(\sqrt{\frac{n\ln n}{\ell_{1,n}k_{1,n}}}\right),$$

and

$$\sup_{k_{1,n}\le k\le k_{2,n}} \sup_{\ell_{1,n}\le \ell\le \ell_{2,n}} |\widehat{r^x}(y) - r^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_2}\right\} + O\left(\frac{\ell_{2,n}}{n}\right)^{\beta_3} + O_{a.co.}\left(\sqrt{\frac{n\ln n}{\ell_{1,n}k_{1,n}}}\right).$$

The proof of this theorem is given in Section 6.

4. Application: automatic data-driven bandwidth selection

Recall that one of the main interests of the kNN method (over usual kernel approaches) is to provide a location adaptive smoothing parameter which is easy to select in practice, because it depends on a single parameter. However, theoretical advances for such a data-driven kNN procedure have still not been stated in the literature and we will show below how the UINN results stated in Section 3 are of interest for this purpose.

We will start by considering the most popular data-driven selection technique which is the leave-one-out cross validation. Then, we will conclude this section by discussing how similar results hold also easily for any other kind of automatic selection rule. In the sequel, when it is not explicitly specified, the minimization over the number of neighbors k (resp. ℓ) has to be understood over the values $k \in (k_{1,n}, k_{2,n})$ (resp. over $j \in (j_{1,n}, j_{2,n})$).

(i) **Regression model:**

The leave-one-out cross-validation procedure [37] consists in minimizing the following squared prediction error criterium:

$$k_{\mathtt{CV}.opt} = \arg \min_{k \in (k_{1,n}, k_{2,n})} \mathtt{CV}(k),$$

where

$$\mathtt{CV}(k) = \sum_{i=1}^{n} \{Y_i - \widehat{m}^{-i}(X_i)\|^2 \quad \text{and} \quad \widehat{m}^{-i}(x) = \frac{\sum_{j \neq i} K\{H_{k,x}^{-1} d(x, X_j)\} Y_j}{\sum_{j \neq i}^{n} K\{H_{k,x}^{-1} d(x, X_j)\}}. \tag{10}$$

The selected value $k_{\mathtt{CV}.opt}$ is a random function which depends on the whole statistical sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ and this makes the direct study of the corresponding kNN estimator rather difficult. The UINN results stated below allow for getting in a very simple way the convergence rate of

$$\widetilde{m}_{\mathtt{CV}} = \frac{\sum_{i=1}^{n} K\{H^{-1}_{k_{\mathtt{CV}.opt}(X_1, Y_1, \ldots, X_n, Y_n), x} d(x, X_j)\} Y_j}{\sum_{i=1}^{n} K\{H^{-1}_{k_{\mathtt{CV}.opt}(X_1, Y_1, \ldots, X_n, Y_n), x} d(x, X_j)\}}.$$

This is stated in the next corollary.

Corollary 4.1. *Under conditions of Theorem 3.1, we have*

$$|\widetilde{m}_{\mathtt{CV}}(x) - m(x)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right).$$

This result is a direct consequence of the first assertion of Theorem 3.1.

(ii) **Conditional distribution function:**

In the multivariate setting, De Gooijer and Gannoun [14] have proposed some cross-validation criterion which can be adapted to the functional setting in the following way:

$$\mathtt{CV}_{GG}(k) = \sum_{i=1}^{n}\{\mathbf{1}_{(Y_i \le Y_j)} - \widehat{F}^{X_i^{-i}}(Y_j)\}^2,$$

where

$$\widehat{F}^{X_i^{-i}}(y) = \frac{\sum_{j\neq i} K\{H_{k,x}^{-1}d(X_i, X_j)\}\mathbf{1}_{\{Y_j\le y\}}}{\sum_{j\neq i}^{n} K\{H_{k,x}^{-1}d(X_i, X_j)\}}.$$

The following corollary ensures the good asymptotic behavior of the data-driven estimator constructed with the optimal number of neighbors which minimizes the $\mathtt{CV}_{GG}$ rule.

Corollary 4.2. *Under the conditions of Theorem 3.1, if $\widehat{F}^x_{CV}(y)$ is the estimator of $F^x(y)$ constructed with the cross-validation procedure $\mathtt{CV}_{GG}$, then we have*

$$|\widehat{F}^x_{CV}(y) - F^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_1}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right).$$

This result is a direct consequence of the second assertion of Theorem 3.1.

(iii) **Conditional probability density:** The problem of the bandwidth selection in conditional density estimation has received some attention in the multivariate statistical literature. The popular cross-validation ideas (see Youndjé et al. [43] and references therein) can be adapted to the functional setting in the following way. The objective is to mininize the following errors:

$$\mathtt{err}_1(\widehat{f}_{(a,b)}, f) = \int\int \left\{\widehat{f}^x_{(a,b)}(y) - f^x(y)\right\}^2 W_1(x)W_2(y)\, dP_X(x)\, dy,$$

$$\mathtt{err}_2(\widehat{f}_{(a,b)}, f) = \frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{f}^{X_i}_{(a,b)}(Y_i) - f^{X_i}(Y_i)\right\}^2 \frac{W_1(X_i)W_2(Y_i)}{f^{X_i}(Y_i)}$$

or

$$\mathtt{err}_3(\widehat{f}, f) = \int\int \mathrm{E}\left\{\widehat{f}^{x}_{(a,b)}(y) - f^{x}(y)\right\}^2 W_1(x)W_2(y)dP_X(x)dy,$$

where W_1 and W_2 are some non-negative weight functions. These theoretical errors are uncomputable in practice and the following leave-one-out cross-validation criterion can be constructed to approximate them in some fully data-driven way:

$$\mathtt{CV}_{CD}(a,b) = \frac{1}{n}\sum_{i=1}^{n} W_1(X_i)\int\left(\widehat{f}^{X_i^{-i}}_{(a,b)}\right)^2(y)W_2(y)dy - \frac{2}{n}\sum_{i=1}^{n}\widehat{f}^{X_i^{-i}}_{(a,b)}(Y_i)W_1(X_i)W_2(Y_i) \quad (11)$$

where

$$\widehat{f}^{X_i^{-i}}_{(a,b)} = b^{-1}\frac{\sum_{j\neq i}^{n} K\{a^{-1}d(X_i,X_j)\}K(b^{-1}|y-Y_j|)}{\sum_{j\neq i}^{n} K\{a^{-1}d(X_i,X_j)\}}.$$

Then the bidimensional smoothing parameter (k,ℓ) is selected by the following procedure:

$$(k_{\mathtt{CV}_{CD}.opt}, \ell_{\mathtt{CV}_{CD}.opt}) = \arg\min_{k_{1,n}\leq k\leq k_{2,n}, \ell_{1,n}\leq \ell\leq \ell_{2,n}} \mathtt{CV}_{CD}(k,\ell).$$

The UINN results, stated in Section 3, allow to obtain the following convergence rate for the data-driven functional cross-validated conditional density operator.

Corollary 4.3. *Under conditions of Theorem 3.1, if $\widehat{f}_{\mathtt{CV}_{CD}}$ is the conditional density estimator constructed with the cross-validation procedure $\mathtt{CV}_{CD}$, then we have*

$$|\widehat{f}^{x}_{\mathtt{CV}_{CD}}(y) - f^{x}(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_2}\right\} + O\left(\frac{\ell_{2,n}}{n}\right)^{\beta_3} + O_{a.co.}\left(\sqrt{\frac{n\ln n}{\ell_{1,n}k_{1,n}}}\right).$$

This result is a direct consequence of the third assertion of Theorem 3.1.

(iv) **Some other data-driven methods:** While the cross-validation procedures described above aim to approximate quadratic errors of estimation, alternative ways for choosing smoothing parameters could be introduced aiming rather to optimize the predictive power of the method. This can be done by minimizing one of the following prediction criterion:

$$(\widetilde{k}_{\mathtt{CV}}^{(1)}, \widetilde{\ell}_{\mathtt{CV}}^{(1)}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i^{(1)})^2, \tag{12}$$

or

$$(\widetilde{k}_{\mathtt{CV}}^{(2)}, \widetilde{\ell}_{\mathtt{CV}}^{(2)}) = \arg \min_{k_{1,n} \leq k \leq k_{2,n}, \ell_{1,n} \leq \ell \leq \ell_{2,n}} \sum_{i=1}^{n} (Y_i - \widehat{Y}_i^{(2)})^2 \tag{13}$$

where the prediction is performed by means either of the conditional median, i.e.,

$$\widehat{Y}_i^{(1)} = \widehat{F}^{(X_i^{-i})^{-1}}(1/2)$$

or by means of the conditional mode, viz.

$$\widehat{Y}_i^{(2)} = \arg\max \widehat{f}^{X_i^{-i}}(y).$$

As far as we know, such selectors did not receive mathematical attention in the past (even in the usual multivariate situation), probably because of technical difficulties. However, the high degree of generality of the UINN results stated above allows us, as an easy direct consequence of Theorem 3.1, to state the following rates of convergence for the corresponding data-driven estimators.

Corollary 4.4. *Let $j = 1$ or $j = 2$, and denote by $\widehat{F}^{(j),x}$, $\widehat{f}^{(j),x}$ and $\widehat{h}^{(j),x}$ respectively the estimators $\widehat{F}$, $\widehat{f}$ and $\widehat{h}$ which are constructed with plugging the optimal numbers of neighbors $\widetilde{k}_{\mathtt{CV}}^{(j)}$ and $\widetilde{l}_{\mathtt{CV}}^{(j)}$. Then, the following three* *results hold:*

$$\begin{aligned}
|\widehat{F}^{(j),x}(y) - F^x(y)| &= O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_1}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right), \\
|\widehat{f}^{(j),x}(y) - f^x(y)| &= O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_2}\right\} + O\left(\frac{\ell_{2,n}}{n}\right)^{\beta_3} + O_{a.co.}\left(\sqrt{\frac{n\ln n}{\ell_{1,n}k_{1,n}}}\right)
\end{aligned}$$

and

$$|\widehat{r}^{(j),x}(y) - r^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_2}\right\} + O\left(\frac{\ell_{2,n}}{n}\right)^{\beta_3} + O_{a.co.}\left(\sqrt{\frac{n\ln n}{\ell_{1,n}k_{1,n}}}\right).$$

5. A simulation study

The objective of this section is two-fold. First, we will show that the automatic kNN procedures can be easily implemented. Then, we will compare them with the usual kernel procedure in order to stress how the local feature of the kNN approach allow to reduce nicely the prediction errors.

5.1. Presentation of the study

We consider the following functional nonparametric model:

$$\forall_{i\in\{1,\ldots,n=300\}} \quad Y_i = m(X_i) + \varepsilon_i, \tag{14}$$

where the ε_i's are generated independently according to a $\mathcal{N}(0, 0.05)$ distribution and are assumed to be independent of X_i for all $i \in \{1, \ldots, n\}$. The sampled functional explanatory variables are generated as follows, for each $i \in \{1, \ldots, n\}$:

$$\forall_{t\in(0,\pi)} \quad X_i(t) = \cos(2a_it) + \sin(2t + b_i) + c_it,$$

where $a_i \sim \mathcal{N}(0, 1)$, $b_i \sim \mathcal{N}(3, 1)$, and $c_i \sim \mathcal{U}(0, 1)$. The curve $X_1, \ldots, X_n$ are then discretized on the same grid generated from 100 equispaced measurements in $(0, \pi)$. The curves are plotted in Figure 1.

The scalar response Y_i is defined by (14) by using the regression operator:

$$m(x) = 5\exp\left[\frac{1}{\displaystyle\int_0^\pi \{1 + x^2(t)\}dt}\right].$$

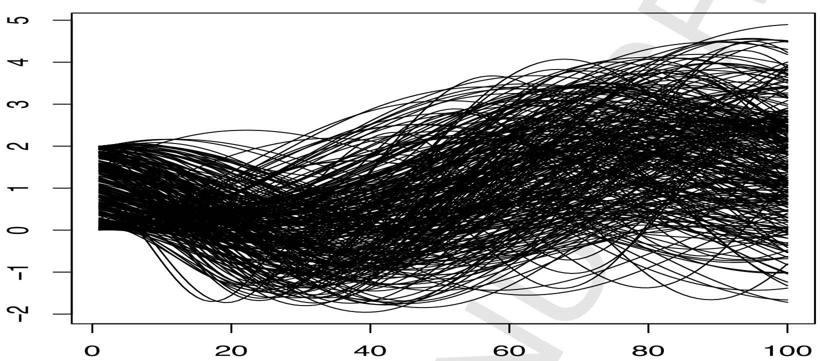


Figure 1: A sample of 300 curves

All along the construction of the various kNN estimators we have used the semi-metric

$$d(u, v) = \int_0^\pi \{u(t) - v(t)\}^2 dt$$

and the kernel function

$$K(t) = \frac{3}{4}\,(1 - t^2)\mathbf{1}_{(0,1)}.$$

The routines used for computing the functional estimators are available at the website `www.lsp.ups-tlse.fr/staph/npfda`.

5.2. Automatic selection of the numbers of neighbors

First of all we will describe how the numbers of neighbors are crucial parameters for the behavior of the estimators. To cover the wide variety of operators (regression, conditional distribution and conditional density operator) we will look at predictors based on these operators (conditional expectation, conditional median and conditional quantile). So, the techniques used for neighbors selection are those motivated in (10), (12) and (13). We have computed the corresponding prediction errors:

$$\mathtt{MSE}_{reg}(k) = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - \widehat{Y}_i^{(reg^k)}\big)^2,$$

$$\texttt{MSE}_{median}(k) = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - \widehat{Y}_i^{(median^k)}\big)^2$$

and

$$\texttt{MSE}_{mode}(k,\ell) = \frac{1}{n}\sum_{i=1}^{n}\big(Y_i - \widehat{Y}_i^{(mode^{k,\ell})}\big)^2$$

where

$$\begin{aligned}\widehat{Y}_i^{(reg^k)} &= \frac{\sum_{j\neq i} K\{H_{k,x}^{-1}d(x,X_j)\}Y_j}{\sum_{j\neq i}^{n} K\{H_{k,x}^{-1}d(x,X_j)\}},\\ \widehat{Y}_i^{(median^k)} &= \widehat{F}^{(X_i^{-i})}(1/2),\end{aligned}$$

and

$$\widehat{Y}_i^{(mode^{k,\ell})} = \arg\max \widehat{f}^{X_i^{-i}}(y).$$

Such errors are evaluated over a sequence of numbers of neighbors defined by $\{5,\ 10,\ 15,\ \ldots,\ 300\}$. The estimator $\widehat{Y}_i^{(reg^k)}$ is obtained by the R-routine named `funopare.`k`NN.gcv`, while the estimators $\widehat{Y}_i^{(median^k)}$ and $\widehat{Y}_i^{(mode^{k,\ell})}$ are obtained by the routines `funopare.mode.lcv` and `funopare.quantile.lcv`.[6].

The results are summarized in Figure 2 for the regression and the conditional median predictors. Concerning the conditional mode, because the estimator involves two parameters ℓ and k, the results are presented in Figure 3 by means of two plots: the left panel shows how the behavior of the error as a function of ℓ (when k is fixed to its optimal value k_{opt}) while the plot in the right panel shows the behavior of the error as function of k (when k is fixed to its optimal value ℓ_{opt}). In each plot a horizontal line shows the minimal value of the prediction error.

The high variability of the `MSE` errors show that the kNN methods are very sensitive to the choice of the number of neighbors, and stress the interest for choosing in each situation the optimal value which is the one obtained by minimizing the above discussed data-driven criterion.

[6] All these routines are available at `www.lsp.ups-tlse.fr/staph/npfda`

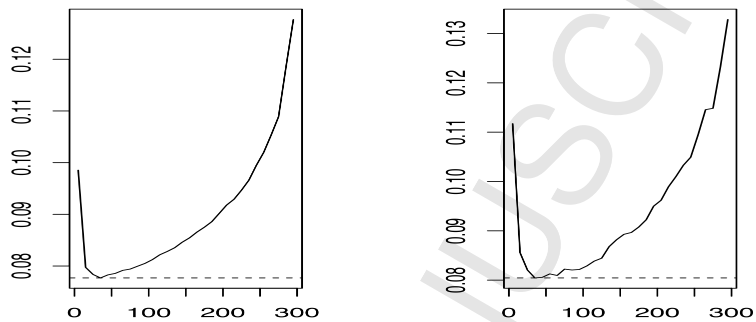


Figure 2: Prediction errors as function of the number of neighbors. Left plot: Regression predictor; Right plot: conditional median predictor.

5.3. Comparative study

The aim of this section is to compare the kNN approach with an usual global kernel one. For that, we have computed the optimal errors with kNN approaches, namely:

$$k\texttt{NNMSE}(reg) := \texttt{MSE}_{reg}(k.opt)$$

$$k\texttt{NNMSE}(median) := \texttt{MSE}_{median}(k.opt)$$

and

$$k\texttt{NNMSE}(mode) := \texttt{MSE}_{mode}(k.opt, \ell.opt)$$

and we compared with their analogous obtained with global fixed bandwidth:

$$\texttt{KERMSE}(reg) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i^{(reg^{a.opt})})^2$$

$$\texttt{KERMSE}(median) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i^{(median^{a.opt})})^2$$

and

$$\texttt{KERMSE}(mode) := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{Y}_i^{(mode^{a.opt,\, b.opt})})^2$$

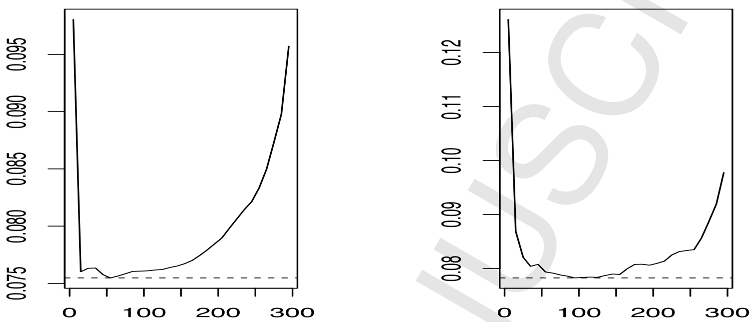


Figure 3: Prediction errors as function of the number of neighbors with conditional mode predictor. Left plot: k is fixed; Right plot: ℓ is fixed.

where

$$\widetilde{Y}_i^{(reg^a)} = \frac{\sum_{j\neq i} K\{a^{-1}d(x,X_j)\}Y_j}{\sum_{j\neq i}^n K\{a^{-1}d(x,X_j)\}}, \quad \widetilde{Y}_i^{(median^a)} = \widetilde{F_a}^{(X_i^{-i})^{-1}}(1/2)$$

and

$$\widetilde{Y}_i^{(mode^{a,b})} = \arg\max \widetilde{f_{a,b}}^{X_i^{-i}}(y)$$

with

$$\widetilde{F_a}^x = \frac{\sum_{i=1}^n K\{a^{-1}d(x,X_i)\}\mathbf{1}(Y_i \leq y)}{\sum_{i=1}^n K\{a^{-1}d(x,X_i)\}}$$

and

$$\widetilde{f_{a,b}}^x(y) = b^{-1}\frac{\sum_{i=1}^n K\{a^{-1}d(x,X_i)\}K(b^{-1}|y-Y_i|)}{\sum_{i=1}^n K\{a^{-1}d(x,X_i)\}}.$$

It should be stressed here that in the three last cases the optimal smoothing parameters $a.opt$ and/or $b.opt$ are chosen among a sequence of quantiles of the distances vector between the functional variable (resp. between the response variables).

All the results are summarized in Table 1, where we observe that the kNN method leads to an important reduction of the prediction errors (at least of 25%) for each predictor (either based on regression, on conditional distribution or on conditional density). On this example the predictor based on the conditional mode seems to be more efficient than the two other ones which are based on the conditional expectation or on the conditional median.

Methods	Mode	Median	Regression
Kernel method	0.081	0.11	0.11
kNN method	0.046	0.087	0.086

Table 1: MSE results

6. Proofs of the results of Sections 3 and 4

In what follows, when there is no confusion, we denote by C or/and C' any generic positive constants. For brievity, we only give full details for the proof concerning the conditional cumulative distribution function estimator. Proofs of the other cases follow along the same lines and are therefore presented in a much more synthetic way, but they may be obtained on request.

- *Proof of Theorem 3.1.* Let us begin by showing (9). For this, we denote:

$$z_n = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_1} + \sqrt{\frac{\ln n}{k_{1,n}}}\right\}$$

and we write for some $\alpha \in (0,1)$,

$$\begin{aligned}\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)| &= \sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)|\mathbf{1}_{\left\{\phi_x^{-1}\left(\alpha\frac{k_{1,n}}{n}\right)\le H_{k,x}\le \phi_x^{-1}\left(\frac{k_{2,n}}{\alpha n}\right)\right\}} \\ &+ \sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)|\mathbf{1}_{\left\{H_{k,x}\notin\left(\phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right),\quad \phi_x^{-1}\left(\frac{k_{2,n}}{\alpha n}\right)\right)\right\}}.\end{aligned}$$

Thus, for all $\epsilon > 0$, we have

$$\begin{aligned}&\Pr\left\{\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)| \ge \epsilon z_n\right\} \\ &\le \Pr\left\{\sup_{k_{1,n}\le k\le k_{2,n}} |\widehat{F}^x(y) - F^x(y)|\mathbf{1}_{\left\{\phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)\le H_{k,x}\le \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right)\right\}} \ge \frac{\epsilon z_n}{2}\right\} \\ &\quad +\Pr\left\{H_{k,x}\notin\left(\phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right), \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right)\right)\right\}.\end{aligned}$$

So to show that

$$\sup_{k_{1,n}\leq k\leq k_{2,n}} |\widehat{F^x}(y) - F^x(y)| = O_{a.co.}(z_n),$$

it suffices to prove the three following results:

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \Pr\left\{H_{k,x} \leq \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)\right\} < \infty, \tag{15}$$

$$\sum_n \sum_{k=k_{1,n}}^{k_{2,n}} \Pr\left\{H_{k,x} \geq \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right)\right\} < \infty, \tag{16}$$

$$\sup_{\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right)\leq h\leq \phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)} |\widetilde{F^x}(y) - F^x(y)| = O_{a.co.}(z_n), \tag{17}$$

where

$$\widetilde{F^x}(y) = \frac{\sum_{i=1}^n K\{h^{-1}d(x,X_i)\}\mathbf{1}_{\{Y_i\leq y\}}}{\sum_{i=1}^n K\{h^{-1}d(x,X_i)\}}.$$

The proof of (17) is not presented here because it follows, step by step, the same argument as for Theorem 3.1 in Kara et al. (2015). Assertions (15) and (16) require much more attention. Thus the latter will be presented with much details. A key tool is the following version of Chernoff's inequality whose proof will be given at the end of this section.

Lemma 6.1. *Let $U_1,\dots,U_n$ be independent Bernoulli random variables with $\Pr(U_i = 1) = p$ for all $i \in \{1,\dots,n\}$. Set $U = X_1 + \cdots + X_n$ and $\mu = pn$. Then, for any $\omega > 0$, we have*

$$\Pr\{U \geq (1+\omega)\mu\} \leq \exp\left\{-\mu \min\left(\omega^2,\omega\right)/4\right\} \tag{18}$$

and if $\omega \in (0,1)$, we have

$$\Pr\{U \leq (1-\omega)\mu\} \leq \exp\left\{-\mu\left(\omega^2/2\right)\right\}. \tag{19}$$

Then, by using Lemma 6.1, we can write

$$\begin{aligned}
\Pr\left\{H_{k,x} \leq \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)\right\} &= \Pr\left\{\sum_{i=1}^n \mathbf{1}_{B\left(x,\phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)\right)} > k\right\}\\
&= \Pr\left\{\sum_{i=1}^n \mathbf{1}_{B\left(x,\phi_x^{-1}\left(\alpha\frac{k_{1,n}}{n}\right)\right)} > \frac{k}{\alpha k_{1,n}}\alpha k_{1,n}\right\}\\
&\leq \exp\left\{-(k-\alpha k_{1,n})/4\right\}.
\end{aligned}$$

Therefore one has,

$$\sum_{k=k_{1,n}}^{k_{2,n}} \Pr\left\{H_{k,x} \le \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)\right\} \le k_{2,n}\exp\left\{-(1-\alpha)k_{1,n}/4\right\} \le n^{1-\{(1-\alpha)/4\}\frac{k_{1,n}}{\ln n}}.$$

In a similar fashion, we get

$$\Pr\left\{H_{k,x} \ge \phi_x^{-1}\left(\frac{k_{2,n}}{\alpha n}\right)\right\} \le \exp\left\{-\frac{(k_{2,n}-\alpha k)^2}{2\alpha k_{2,n}}\right\}.$$

It follows that

$$\begin{aligned}\sum_{k=k_{1,n}}^{k_{2,n}} \Pr\left\{H_{k,x} \ge \phi_x^{-1}\left(\frac{k_{2,n}}{\alpha n}\right)\right\} &\le k_{2,n}\exp\left\{-(1-\alpha)k_{1,n}/2\alpha\right\} \\ &\le n^{1-\{(1-\alpha)/2\alpha\}\frac{k_{2,n}}{\ln n}}.\end{aligned}$$

Because $k_{1,n}/\ln n \to \infty$ we finally obtain (15) and (16). The proof of (9) is now complete. The other assertions of Theorem 3.1 can be obtained in a similar way.

- *Proof of Corollary 4.2.* Let $\tilde{k}, \tilde{\ell}$ be any random pair taking values into $(k_1, k_2) \times (\ell_1, \ell_2)$, and denote by $\widehat{F}^x$ the estimator, defined in (4), but by using the random numbers of neighbors $\tilde{k}$ and $\tilde{\ell}$. As a direct consequence of the UINN result (9), we get

$$|\widehat{F}^x(y) - F^x(y)| = O\left\{\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\beta_1}\right\} + O_{a.co.}\left(\sqrt{\frac{\ln n}{k_{1,n}}}\right). \qquad (20)$$

It suffices to remark that Corollary 4.2 is a special case of (20) when the pair $(\tilde{k}, \tilde{\ell})$ is the cross-validated one which is obtained by minimizing the criterion $\mathtt{CV}_{GG}$.

- *Proofs of Corollaries 4.1, 4.3 and 4.4.* In the same way, one can obtain results of the same kind as (20) for other kNN estimators using a random number of neighbors, from which Corollaries 4.1, 4.3 and 4.4 follow directly.

- *Proof of the Chernoff-type inequalities in Lemma 6.1.* Let $0 \le \tau \le 1-p$. By the Markov inequality we get, for any $\zeta > 0$,

$$\Pr\left\{U \ge n(p+\tau)\right\} \le \frac{\mathrm{E}(e^{\zeta U})}{e^{\zeta n(p+\tau)}} \le \frac{(pe^{\zeta}+1-p)^n}{e^{\zeta n(p+\tau)}}.$$

Taking now $\zeta = \ln\{(1-p)(p+\tau)/(p(1-p-\tau))\}$, we arrive at the most usual form of Chernoff's inequality, viz.

$$\forall_{\tau\in[0,1-p]} \ \Pr\{U \geq n(p+\tau)\} \leq \Big(\frac{p}{p+\tau}\Big)^{n(p+\tau)} \Big(\frac{1-p}{1-p-\tau}\Big)^{n(1-p-\tau)}. \quad (21)$$

This inequality will be the starting point for proving the two results (18) and (19) stated in Lemma 6.1.

– *Proof of* (18). Let $\omega > 0$. Applying the result (21) with $\tau = p\omega$ leads to

$$\Pr\{U \geq \mu(1+\omega)\} \leq \Big(\frac{1}{1+\omega}\Big)^{n\mu(1+\omega)} \Big(\frac{1-p}{1-p-p\omega}\Big)^{n(1-p-p\omega)}. \quad (22)$$

By a simple two-order Taylor expansion, we can also show that

$$\ln\left\{(1+\omega)^{\mu(1+\omega)}\left(\frac{1-p-p\omega}{1-p}\right)^{1-p-p\omega}\right\} \geq \frac{p}{4}\min(\omega^2,\omega). \quad (23)$$

Finally, the claimed result (18) follows directly from (22) and (23).

– *Proof of* (19). By applying the result (21) to the variables $V_i = -U_i$ and by taking $\tau = p\omega$, we get directly that

$$\Pr\{U \leq (1-\omega)\mu\} \leq \frac{e^{-\mu\omega}}{(1-\omega)^{\mu(1-\omega)}}. \quad (24)$$

Recall that to get the assertion (19), we assumed that $0 < \omega < 1$. By some simple analytic arguments one can see that

$$\ln(1-\omega) \geq \frac{\omega^2/2-\omega}{1-\omega}. \quad (25)$$

Finally, the claimed result (19) follows directly from (24) and (25).

7. Concluding remarks

This paper has started by providing UINN (uniform in number of neighbors) rates of consistency for a wide class of nonparametric kNN estimators with functional data (see Section 3) contributing to the knowledge of kNN estimators of the functional regression. The main use of this new kind of asymptotic results is to derive asymptotic theory for data-driven kNN estimators (see Section 4). The paper is written in a rather general form allowing for several different operators estimation problems (regression, density, cumulative distribution function, and hazard operators), and for a wide scope of data-driven number of neighbors selectors. A special attention is given, in Section 4, to selection procedures based on cross-validation ideas. The results of Section 4 are, as far as we know, the first ones in the literature concerning operators other than regression. Moreover, and maybe more importantly, they are the first ones in the functional kNN setting. We would like to stress that our paper has been deliberately theoretically oriented. However, in spite of the technical feature linked with UINN asymptotics, the mathematical results provided in Section 3 are not only interesting in themselves but also (and maybe more importantly) for their direct applied impacts since the consequences derived in Section 4 make the kNN approaches automatically usable in practice. These facts are briefly mentioned in the simulations presented in Section 5.

It should be stressed that even if functional data is the main purpose of this work, it has been written in such a way that it also be applied directly to the multivariate setting; it suffices to take $\mathcal{F} = \mathbb{R}^p$ in the models of Section 2. In such a specific multivariate situation, as well as the UINN results of Section 3, the data-driven rules of Section 4 are extensions of those existing in the literature.

To conclude, it is important to keep in mind that kernel techniques are not also useful in purely nonparametric models like those investigated herein but also in semi-parametric models. The study of data-driven kNN methods in the functional semi-parametric modeling is, as far as we known, a totally underdeveloped field. Our guess is that the UINN approach developed in our paper could also be used in many semi-parametric functional situations such as the functional projection pursuit regression [10], single index model [23], partial linear models [1, 33] or sparse modeling [2].

Acknowledgements.. The authors would like to thank the four anonymous reviewers and specially the Editor-in-chef of JMVA, Prof. C. Genest, for their valuable comments and suggestions which improved substantially the quality of an earlier version of this paper.

References

[1] Aneiros G. and Vieu, P. (2006). Semi-functional partial linear regression. *Statist. Probab. Lett.*, **76**, 1102–1110.

[2] Aneiros G. and Vieu, P. (2015). Partial linear modelling with multi-functional covariates. *Comput. Statist.*, **30**, 647–671.

[3] Beirlant, J., Berlinet, A. and Biau, G. (2008). Higher order estimation at Lebesgue points. *Ann. Inst. Statist. Math.*, **60**, 651–677.

[4] Benhenni, K., Ferraty, F., Rachdi, M. and Vieu, P. (2007). Local smoothing regression with functional data. *Comput. Statist.*, **22**, 353–369.

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

[6] Bongiorno E., Salinelli, E., Goia, A. and Vieu, P. (Eds.) (2014). An overview of IWFOS'2014. In *Contributions in Infinite-Dimensional Statistics and Related Topics*, 1–6, Esculapio, Bologna.

[7] Burba, F., Ferraty, F. and Vieu, P. (2009). k-nearest neighbor method in functional nonparametric regression. *J. Nonparametr. Statist.*, **21**, 453–469.

[8] Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Statist.*, **10**, 340–355.

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

[10] Chen, D., Hall, P., Müller, H.G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.*, **38**, 3458–3486.

[11] Collomb, G. (1981). Estimation non paramétrique de la régression: Revue bibliographique. *Internat. Statist. Rev.*, **49**, 75–93.

[12] Cover, T.M. (1968). Estimation by the nearest neighbor rule. *IEEE Trans. Inform. Theory*, **IT-14**, 50–55.

[13] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *J. Statist. Plann. Inference,* **147**, 1–23.

[14] De Gooijer, J. and Gannoun, A. (2000). Nonparametric conditional predictive regions for time series. *Comput. Statist. Data Anal.*, **33**, 259–257.

[15] Deheuvels, P. and Ouadah, S. (2013). Uniform-in-bandwidth functional limit laws. *J. Theoret. Probab.*, **26**, 697–721.

[16] Devroye, L. Györfi, L. Krzyzak, A. and Lugosi, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *Ann. Statist.*, **22**, 1371–1385.

[17] Devroye, L. and Wagner, T. (1977). The strong uniform consistency of nearest neighbor density. *Ann. Statist.*, **5**, 536–540.

[18] Devroye, L. and Wagner, T. (1982). Nearest neighbor methods in discrimination. In *Classification, pattern recognition and reduction of dimensionality*, 193–197, Handbook of Statistics, **2**, North-Holland, Amsterdam.

[19] Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, **33**, 1380–1403.

[20] Ezzahrioui, M., Ould-Saïd, E. (2008). Asymptotic normality of nonparametric estimator of the conditional mode for functional data. *J. Nonparametr. Statist.*, **20**, 3–18.

[21] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice.* Springer, New York, 2006.

[22] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. (2010). Rate of uniform consistency for nonparametric estimates with functional variables. *J. Statist. Plann. Inference*, **140**, 335–352.

[23] Goia, A. and Vieu, P. (2015). A partitioned single functional index model. *Comput. Statist.*, **30**, 676–692.

[24] Goia, A. and Vieu, P. (2016). An introduction to recent advances in high/infinite dimensional statistics. *J. Multivariate Anal.*, **146**, 1–6.

[25] Györfi, L., Kohler, Krzyzak, A. and Walk, H. (2002). *A Distribution-Free theory of Nonparametric Regression.* Springer, New York.

[26] Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data With Applications.* Springer, New York.

[27] Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators.* Wiley, Chichester.

[28] Kara, L.Z., Laksaci, A., Rachdi, M. Vieu, P. (2016). Uniform in bandwidth consistency for various kernel estimators involving functional data. *J. Nonparameter. Statist.*, In press.

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

[30] Laloë, T. (2008). A k-nearest approach for functional regression. *Statist. Probab. Lett.*, **10**, 1189–1193.

[31] Li, J.P. (1985). Strong convergence rates of error probability estimation in the nearest neighbor discrimination rule. *J. Math. (Wuhan)*, **5**, 113–118.

[32] Lian, H. (2011a). Convergence of functional k-nearest neighbor regression estimate with functional responses. *Electron. J. Statist.*, **5**, 31–40.

[33] Lian, H. (2011b). Functional partial linear model. *J. Nonparametr. Statist.*, **23**,115–128.

[34] Masry, E. (2005). Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stoch. Proc. and their Appl.,* **115**, 155–177.

[35] Moore, D. and Yackel, J. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.*, **5**, 143–154.

[36] Rachdi, M., Laksaci, A., Demongeot, J., Abdali, A. and Madani, F. (2014). 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.

[37] Rachdi, M. and Vieu, P. (2007). Nonparametric regression for functional data: automatic smoothing parameter selection. *J. Statist. Plann. Inference,* **137**, 2784–2801.

[38] Ramsay, J.O. and Silverman, B. (2005). *Functional Data Analysis.* Springer, New York.

[39] Scott, D. (2015). *Multivariate Density Estimation (2nd Ed)*. Wiley, New York.

[40] Stone, C.J. (1977). Consistent nonparametric regression (with discussion). *Ann. Statist.*, **5**, 595–645.

[41] Tran, T., Wehrens, R. and Buydens, L. (2006). kNN-kernel density-based clustering for high-dimensional multivariate data. *Comput. Statist. Data Anal.*, **51**, 513–525.

[42] Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.

[43] Youndjé, E., Sarda, P. and Vieu, P. (1993). Kernel estimator of conditional density: bandwidth selection for dependent data. *C. R. Acad. Sci. Math. Paris*, **316**, 935–938.

[44] Zhang, J. (2013). *Analysis of Variance for Functional Data*. Chapman & Hall/CRC, London.