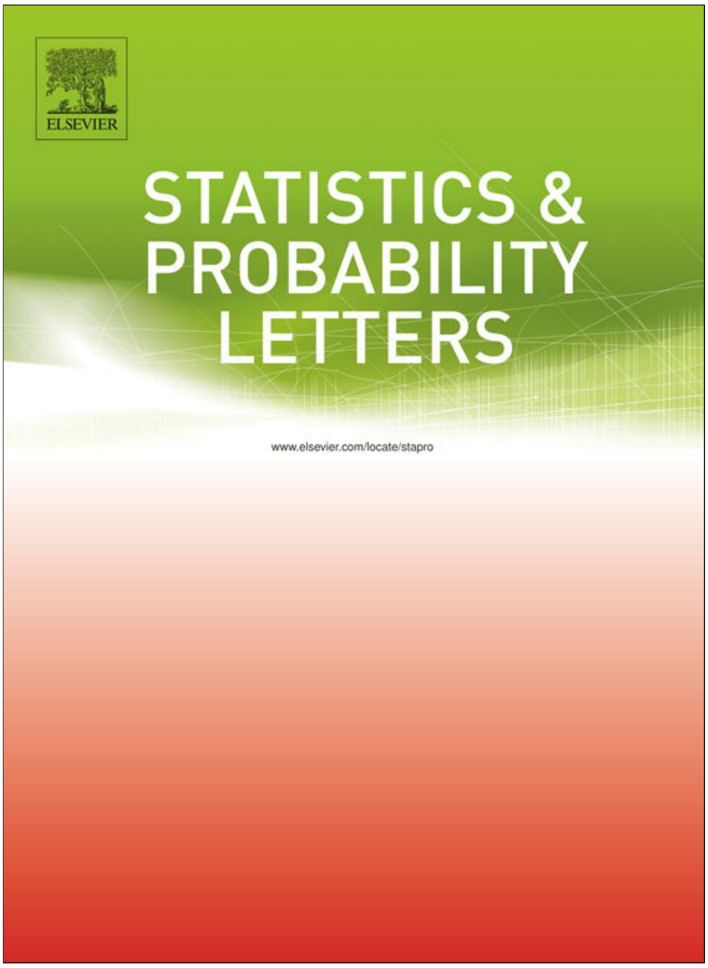
ELSEVIER
STATISTICS &
PROBABILITY
LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 83 (2013) 1863–1870

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Uniform consistency of kNN regressors for functional variables

Nadia L. Kudraszow [a,*], Philippe Vieu [b]

[a] CONICET, Universidad de La Plata, Argentina

[b] Institut de Mathématiques, Université Paul Sabatier, Toulouse, France

ARTICLE INFO

Article history:
Received 2 March 2012
Received in revised form 17 April 2013
Accepted 17 April 2013
Available online 24 April 2013

Keywords:
Functional data analysis
kNN regression
Nonparametric statistics
Uniform consistency

ABSTRACT

This paper is devoted to nonparametric analysis of functional data. We give asymptotic results for a kNN generalized regression estimator when the observed variables take values in any abstract space. The main novelty is our uniform consistency result (with rates).

1. Introduction

There is a wide variety of nonparametric smoothing techniques which are constructed by means of different sorts of local weighting. This is the case for kernel regression methods where the regression estimate is constructed, at each point of interest x, by averaging the data falling into a neighbourhood of x. It is well-known that the key parameter for insuring the good behaviour of the estimate (both from an asymptotic point of view and for finite sample applications) is the size of the neighbouring window, namely the radius h of the ball defining it. On one side, there is the wish for letting this window width depend on x (that is, $h = h_x$), in order to construct location-adaptive smoothers. On the other hand, the choice of h_x in practice is not straightforward. One of the most popular ways of overcoming these difficulties is to consider a kNN neighbourhood. It consists in fixing some integer k and in defining h_x as a value such that the ball of centre x and radius h_x contains exactly k data elements. In this way, the resulting estimate is location-adaptive and it depends on a single discrete parameter (that is, the integer k) instead of depending on an infinite number of continuous parameters (those are, the radius $h = h_x$, for all x). As a partial drawback, the mathematical treatment of kNN estimates is much more difficult because the method is based on a random neighbourhood. In situations where data are real valued the literature is rather large and kNN regressors are known to combine good theoretical asymptotic properties and a nice finite sample behaviour (see Collomb (1980) or Devroye (1982), for early references, and Chapter 6 in Györfi et al. (2002), for a general discussion). Extensions to multivariate data have also been widely studied (see for instance Bhattacharya and Mack (1990), for an early reference, and Biau and Devroye (2010), most recent advances).

On the other hand, modern sciences are now able to collect continuous data (such as curves and images) and statisticians are in front of new challenges. This field, known as functional data analysis (FDA), has been popularized in the last two decades, mainly through the books by Ramsay and Silverman (2002, 2005), which are devoted to methodological and applied

* Corresponding author. Tel.: +54 91155607451.
E-mail addresses: nkudraszow@mate.unlp.ar, n_kudraszow@yahoo.com (N.L. Kudraszow), philippe.vieu@math.univ-toulouse.fr (P. Vieu).

0167-7152/$ – see front matter
http://dx.doi.org/10.1016/j.spl.2013.04.017

issues respectively. While this early literature was mainly based on linear ideas, the link between nonparametric modelling and FDA started recently to be studied with the book by Ferraty and Vieu (2006). Nonparametric functional data analysis (NPFDA) is now taking more and more importance in the literature, and specially in the context of regression analysis with functional covariates (see for instance Ferraty and Romain (2010)). Because the local features of a random variable change more from one area of the space to another one when the dimension increases, kNN methods are natural appealing candidates for nonparametric regression in infinite dimensional settings. Until now the literature in question has been quite underdeveloped since it is restricted to pointwise consistency results (see Burba et al. (2009) and Lian (2011)).

The main aim of our paper is to state a general consistency result on kNN regression estimates with functional variables (see Theorem 2 in Section 3). The asymptotics are stated in terms of almost complete convergence[1] which is known to imply both almost sure convergence and convergence in probability (see for instance Bosq and Lecoutre (1987)). The main feature of our result is uniformity, and it is discussed in Section 5 how this is of utmost importance for many further developments in FDA. The second feature is to provide the rates of convergence which show the effects of the topological structure endowing the functional space. It is finally worth noting that, even if our main purpose is functional regression, the paper is written in a general form allowing other types of functional estimation problems (see Remark 1). Also, as a second interesting by-product, we state a technical lemma showing how uniform results for local weighted sequences on fixed neighbourhood can be adapted to similar results with random neighbourhoods. The interest of such a result is well beyond the scope of this paper (see Section 4.1).

2. The kNN functional regressor

2.1. General notations

Let $\{(\mathbf{X}_i, Y_i)\}_{i=1,\ldots,n}$ be n random independent and identically distributed (i.i.d.) realizations of $(\mathbf{X}, Y)$ and valued in $\mathcal{F} \times \mathbb{R}$. $(\mathcal{F}, d)$ is a semi-metric space; $\mathcal{F}$ is not necessarily of finite dimension and we do not suppose the existence of some dominating measure for the probability distribution of $\mathbf{X}$. For $\chi \in \mathcal{F}$, let $B(\chi, \varepsilon)$ be the ball of centre χ and radius ε for the topology associated with d: $B(\chi, \varepsilon) = \{\chi' \in \mathcal{F} / d(\chi', \chi) \leq \varepsilon\}$.

Our purpose is to prove uniform results on a subset $S_{\mathcal{F}}$ of $\mathcal{F}$ whose topological structure appears through the notion of entropy. Recall that the Kolmogorov's ε-entropy of some set S is defined by $\psi_S(\varepsilon) = \log(N_\varepsilon(S))$, where $N_\varepsilon(S)$ is the minimal number of open balls in $\mathcal{F}$ of radius ε which is necessary to cover S.

2.2. The model and the estimate

We wish to estimate a generalized regression function, defined from some known real-valued Borel link L, as

$$m_L(\chi) = E[L(Y)|\mathbf{X} = \chi], \quad \forall \chi \in \mathcal{F}. \tag{2.1}$$

Remark 1. It is worth noting that this model includes the standard regression problems (when L is the function $L(t) = t$), but also many other ones. For instance when choosing, for fixed y, the function $L(t) = 1_{(-\infty,y]}(t)$ one gets as operator m_L the conditional distribution operator $F_y(\chi) = P(Y \leq y|\mathbf{X} = \chi)$. Similarly, the choice $L(t) = 1_{(y,\infty)}(t)$ leads to the conditional survival function.

The kNN kernel estimator is defined, for any functional element χ, by looking for the k nearest neighbours of χ among the functional sample $\{\mathbf{X}_i,\ i = 1, \ldots, n\}$, and then by a weighted averaging of the corresponding $L(Y_i)$. This leads to

$$\widehat{m}_{k\mathrm{NN}}(\chi) = \frac{\sum_{i=1}^{n} K\left(H_{n,k}(\chi)^{-1} d(\chi, \mathbf{X}_i)\right) L(Y_i)}{\sum_{i=1}^{n} K\left(H_{n,k}(\chi)^{-1} d(\chi, \mathbf{X}_i)\right)}, \quad \forall \chi \in \mathcal{F}, \tag{2.2}$$

where K is a kernel with support in $[0, \infty)$ and $H_{n,k}(\chi)$ is defined as follows:

$$H_{n,k}(\chi) = \min\left\{h \in \mathbb{R}^+ : \sum_{i=1}^{n} 1_{B(\chi,h)}(\mathbf{X}_i) = k\right\}. \tag{2.3}$$

Note that $H_{n,k}(\chi)$ is a positive random variable which depends on $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$.

[1] Recall that a sequence $(T_n)_{n\in\mathbb{N}^*}$ of random variables is said to converge almost completely to some variable T, if and only if $\forall\varepsilon > 0$, $\sum_{n=1}^{\infty} P(|T_n - T| > \varepsilon) < \infty$. This is denoted by $T_n \to T$, a.co. (or equivalently by $T_n - T = o_{\text{a.co.}}(1)$). We say the sequence $(T_n)_{n\in\mathbb{N}^*}$ is $T_n - T = o_{\text{a.co.}}(u_n)$, if the sequence $u_n^{-1}|T_n - T|$ converges almost completely to zero.

In fact this kNN estimate can be seen as an extension to random and locally adaptive neighbourhood of the traditional kernel approach defined as

$$\widehat{m}_L(\chi) = \frac{\sum_{i=1}^{n} K\left(h(\chi)^{-1}d(\chi, \mathbf{X}_i)\right) L(Y_i)}{\sum_{i=1}^{n} K\left(h(\chi)^{-1}d(\chi, \mathbf{X}_i)\right)}, \quad \forall \chi \in \mathcal{F}, \tag{2.4}$$

where $h(\chi) = h_n(\chi)$ are positive real numbers decreasing to zero as n goes to infinity. The principal interest of the kNN method appears in the implementation. The fact that the smoothing parameter k takes its values in a discrete set makes things more simple from the practical point of view. Burba et al. (2009) showed by means of examples that the kNN method takes into account the local structure of the data and gives better predictions when the data are heterogeneously concentrated.

3. Rates of uniform consistency

3.1. The hypotheses

We denote by C and C' some generic strictly positive real constants, changing from line to line. Assume the following hypotheses.

(H1) (H1a) $\forall \varepsilon > 0$, $\varphi_\chi(\varepsilon) := P(\mathbf{X} \in B(\chi, \varepsilon)) > 0$, with $\varphi_\chi(\cdot)$ continuous on a neighbourhood of 0 and $\varphi_\chi(0) = 0$.
(H1b) There exist a nonnegative function $\phi(.)$ and a positive function $f(.)$ such that for some $b > 0$, $\sup_{\chi \in S_{\mathcal{F}}} \left|\varphi_\chi(\varepsilon)/\phi(\varepsilon) - f(\chi)\right| = O(\varepsilon^b)$, as $\varepsilon \to 0$.
(H2) m_L is a bounded Lipschitz operator of order b on $S_{\mathcal{F}}$, that is, there exists $b > 0$ such that $\forall \chi_1, \chi_2 \in S_{\mathcal{F}}$, $|m_L(\chi_1) - m_L(\chi_2)| \le Cd^b(\chi_1, \chi_2)$.
(H3) $\forall l \ge 2$, $E(|L(Y)|^l|\mathbf{X} = \chi) < \delta_l(\chi) < C$ with $\delta_l(\cdot)$ continuous on $S_{\mathcal{F}}$.
(H4) The kernel function K has to be
(H4a) a nonnegative, bounded, non increasing function with support [0, 1] and Lipschitz on [0, 1),
(H4b) if $K(1) = 0$, it must also be such that $-\infty < C < K'(t) < C' < 0$.
(H5) The function ϕ is such that
(H5a) $\phi(0) = 0$ and $\lim_{\varepsilon\to 0} \phi(\varepsilon) = 0$,
(H5b) $\exists C > 0, \exists \eta_0 > 0, \forall 0 < \eta < \eta_0, \phi'(\eta) < C$,
(H5c) ϕ is regularly varying at 0 with nonnegative index, that is, there exists a function $\zeta_0(u) = u^\alpha$, with $\alpha \ge 0$, such that for all $u \in [0, 1]$,

$$\lim_{\varepsilon\to 0} \frac{\phi(u\varepsilon)}{\phi(\varepsilon)} := \lim_{\varepsilon\to 0} \zeta_\varepsilon(u) = \zeta_0(u).$$

(H6) Kolmogorov's ε-entropy of $S_{\mathcal{F}}$ satisfies

$$\sum_{n=1}^{\infty} \exp\left\{(1-\omega)\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)\right\} < \infty \quad \text{for some } \omega > 1.$$

Remark 2. In the case of $K(1) = 0$, under assumptions (H1), (H4), (H5a) and (H5b) it is easily seen that $\exists C > 0$, $\exists \eta_0 > 0$, $\forall 0 < \eta < \eta_0$, $\int_0^\eta \phi(u)du > C\eta\phi(\eta)$.

3.2. Main result

We start by reminding the uniform asymptotic properties of $\widehat{m}_L$ defined in (2.4). Theorem 1 was proved by Ferraty et al. (2010) in the special case when $h(\chi) = h$ for all $\chi \in S_{\mathcal{F}}$, but their proof can be followed line by line under (3.1). This general condition (3.1) will be a crucial preliminary tool for us.

Theorem 1. *Suppose that assumptions* (H1)–(H6) *hold, and that the bandwidths $h(\chi)$ satisfy, for n large enough,*

$$0 < C_1 h \le \inf_{\chi \in S_{\mathcal{F}}} h(\chi) \le \sup_{\chi \in S_{\mathcal{F}}} h(\chi) \le C_2 h < \infty, \tag{3.1}$$

where $h = h_n$ is a sequence (independent of χ) such that $h \to 0$ and $(\log n)^2/n\phi(h) < \psi_{S_{\mathcal{F}}}(\log n/n) < n\phi(h)/\log n$, for n large enough. Then we have

$$\sup_{\chi \in S_{\mathcal{F}}} |\widehat{m}_L(\chi) - m_L(\chi)| = O(h^b) + O_{a.co.}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right).$$

We can now state our main result, whose proof will be presented in Section 4.

Theorem 2. *Under assumptions* (H1)–(H6), *suppose that* $k = k_n$ *is a sequence of positive real numbers such that* $k/n \to 0$ *and* $(\log n)^2/k < \psi_{S_\mathcal{F}}(\log n/n) < k/\log n$, *for* n *large enough, then we have*

$$\sup_{\chi\in S_\mathcal{F}} |\widehat{m}_{kNN}(\chi) - m_L(\chi)| = O_{a.co.}\left(\phi^{-1}\left(\frac{k}{n}\right)^b + \sqrt{\frac{\psi_{S_\mathcal{F}}\left(\frac{\log n}{n}\right)}{k}}\right).$$

3.3. Comments on the results

On the hypotheses. First of all note that condition (H4) includes two types of kernels which are traditionally used in practice: box and continuous kernels. Conditions (H2) and (H3) are very common in the nonparametric literature (and also in the multivariate one) and they are used here to prove Theorem 1 (see Ferraty et al. (2010) for details).

Assumption (H1b) is known as the "concentration property" in infinite dimensional spaces. For many examples, the small ball probability $\varphi_\chi(\varepsilon)$ can be approximated, around zero, as a product of two independent functions $f(\chi)$ and $\phi(\varepsilon)$ (see for instance Mayer-Wolf and Zeitouni (1993) for the diffusion process, Bogachev (1999) for a Gaussian measure and Li and Shao (2001) for a general Gaussian process). The most frequent result available in the literature is of the form $\varphi_\chi(\varepsilon) \sim f(\chi)\phi(\varepsilon)$ where $\phi(\varepsilon) = \varepsilon^\gamma \exp(-C/\varepsilon^p)$ with $\gamma \geq 0$ and $p \geq 0$. It corresponds to the Ornstein–Uhlenbeck and general diffusion processes (for such processes, $p = 2$ and $\gamma = 0$) and the fractal processes (for such processes, $\gamma > 0$ and $p = 0$). This class of processes also satisfy condition (H5). Note that these notions are strongly linked with the measure of proximity d which is considered, and all the examples discussed just before are concerned with d being standard norms (like Hölder norm or supremum norm for instance).

Assumption (H6) acts on Kolmogorov's ε-entropy of $S_\mathcal{F}$. There are special cases of functional spaces $\mathcal{F}$ and subsets $S_\mathcal{F}$ where $\psi_{S_\mathcal{F}}(\log(n)/n) \gg \log(n)$. Some examples are the closed ball in a Sobolev space, the unit ball of the Cameron–Martin space and a compact subset in a Hilbert space with a projection semi-metric (see Kolmogorov and Tikhomirov (1959), van der Vaart and van Zanten (2007)and Ferraty et al. (2010), respectively, for further details). In all these cases it is easy to see that (H6) is verified as soon as $\beta > 2$.

On the rates of convergence. First of all it is worth noting that, by taking k of order $n\phi(h)$, the kNN estimate reaches the same rate of convergence as the kernel estimate does (see Theorem 1). More importantly, to attest the quality of these rates, it suffices to look at the case $\mathcal{F} = \mathbb{R}^q$ to see that they are exactly matching the rate $(\log n/n)^{b/(2b+q)}$ which is optimal for q-dimensional functions (see Stone (1982)). Note also that, for the exponential-type processes described before the rate of convergence may look quite slow for unfamiliar people (of order $(\log n)^{-\alpha}$ for some $\alpha > 0$) but this is true only when using as "d" a standard norm; other kinds of d can be used to improve strongly these rates, as discussed in Ferraty and Vieu (2006, Lemma 13.6).

4. Proofs

4.1. A general lemma

This section presents a result that will allow us to derive asymptotic results for kNN estimators directly from similar results on kernel regression. Because this lemma is of possible interest for many other purposes (see discussion in Section 5), this section is self-contained and specific notations are introduced. Let $(\mathbf{A}_i, B_i)_{i=1,\ldots,n}$ be n random pairs valued in $(\Omega\times\mathbb{R}, \mathcal{A}\times\mathcal{B}(\mathbb{R}))$ where $(\Omega, \mathcal{A})$ is a general measurable space. Let S_Ω be a fixed subset of Ω and we note that $G : \mathbb{R}\times(S_\Omega\times\Omega) \to \mathbb{R}^+$ a function such that, $\forall\chi \in S_\Omega, G(\cdot, (\chi, \cdot))$ is measurable and $\forall t, t' \in \mathbb{R}$:

$$(L_0) : t \leq t' \Longrightarrow G(t, \mathbf{z}) \leq G(t', \mathbf{z}) \quad \forall \mathbf{z} \in S_\Omega \times \Omega.$$

Let $(D_n(\chi))_{n\in\mathbb{N}}$ be a sequence of random real variables (r.r.v.) and $c : S_\Omega \to \mathbb{R}$ be a nonrandom function such that $\sup_{\chi\in S_\Omega} |c(\chi)| < \infty$. Also, for all χ in S_Ω and $n \in \mathbb{N}\backslash\{0\}$ we define

$$c_{n,\chi}(t) = \frac{\sum_{i=1}^n B_i G(t, (\chi, \mathbf{A}_i))}{\sum_{i=1}^n G(t, (\chi, \mathbf{A}_i))}.$$

Lemma 3. *Let* $(u_n)_{n\in\mathbb{N}}$ *be a decreasing positive sequence such that* $\lim_{n\to\infty} u_n = 0$. *If, for all increasing sequence* $\beta_n \in (0, 1)$ *with* $\beta_n - 1 = O(u_n)$, *there exist two sequences of r.r.v.* $(D_n^-(\beta_n, \chi))_{n\in\mathbb{N}}$ *and* $(D_n^+(\beta_n, \chi))_{n\in\mathbb{N}}$ *such that*

(L_1) $\forall n \in \mathbb{N}, \forall\chi \in S_\Omega, D_n^-(\beta_n, \chi) \leq D_n^+(\beta_n, \chi)$,
(L_2) $\mathbb{1}_{\{D_n^-(\beta_n,\chi)\leq D_n(\chi)\leq D_n^+(\beta_n,\chi),\ \forall\chi\in S_\Omega\}} \longrightarrow 1$, a.co.

(L_3) $\sup_{\chi\in S_\Omega}\left|\frac{\sum_{i=1}^n G\left(D_n^-(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)}{\sum_{i=1}^n G\left(D_n^+(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)}-\beta_n\right| = O_{\text{a.co.}}(u_n)$,

(L_4) $\sup_{\chi\in S_\Omega}|c_{n,\chi}(D_n^-(\beta_n,\chi))-c(\chi)| = O_{\text{a.co.}}(u_n)$,

(L_5) $\sup_{\chi\in S_\Omega}|c_{n,\chi}(D_n^+(\beta_n,\chi))-c(\chi)| = O_{\text{a.co.}}(u_n)$.

Then

$$\sup_{\chi\in S_\Omega}|c_{n,\chi}(D_n(\chi))-c(\chi)| = O_{\text{a.co.}}(u_n). \tag{4.1}$$

Proof. For technical reasons, we assume in this proof that the random variables B_i are nonnegative. The result for any real-valued B_i can be deduced by taking $B_i = B_i^+ - B_i^-$ where $B_i^+ = \max(B_i, 0)$ and $B_i^- = -\min(B_i, 0)$. Let

$$c_{n,\chi}^-(\beta_n) = \frac{\sum_{i=1}^n B_i G\left(D_n^-(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)}{\sum_{i=1}^n G\left(D_n^+(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)},$$

$$c_{n,\chi}^+(\beta_n) = \frac{\sum_{i=1}^n B_i G\left(D_n^+(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)}{\sum_{i=1}^n G\left(D_n^-(\beta_n,\chi),(\chi,\mathbf{A}_i)\right)}.$$

For all sequence $\beta_n\in(0,1)$ with $\beta_n-1=O(u_n)$, (L_3) and (L_4) give

$$\begin{aligned}\sup_{\chi\in S_\Omega}|c_{n,\chi}^-(\beta_n)-c(\chi)| &\le \sup_{\chi\in S_\Omega}|c_{n,\chi}^-(\beta_n)-\beta_n c(\chi)| + |c(\chi)|\,|\beta_n-1| \\ &= O_{\text{a.co.}}(u_n)\end{aligned} \tag{4.2}$$

while (L_3) and (L_5) give similarly

$$\sup_{\chi\in S_\Omega}|c_{n,\chi}^+(\beta_n)-c(\chi)| = O_{\text{a.co.}}(u_n). \tag{4.3}$$

For all $\varepsilon>0$, we note $T_n(\varepsilon) = \{\sup_{\chi\in S_\Omega}|c_{n,\chi}(D_n(\chi))-c(\chi)|\le\varepsilon u_n\}$ and for all sequence $\beta_n\in(0,1)$ with $\beta_n-1=O(u_n)$

$$S_n^-(\varepsilon,\beta_n) = \left\{\sup_{\chi\in S_\Omega}|c_{n,\chi}^-(\beta_n)-c(\chi)|\le\varepsilon u_n\right\},$$

$$S_n^+(\varepsilon,\beta_n) = \left\{\sup_{\chi\in S_\Omega}|c_{n,\chi}^+(\beta_n)-c(\chi)|\le\varepsilon u_n\right\},$$

$$S_n(\beta_n) = \{c_{n,\chi}^-(\beta_n)\le c_{n,\chi}(D_n(\chi))\le c_{n,\chi}^+(\beta_n), \forall\chi\in S_\Omega\}.$$

Then for all $\beta_n\in(0,1)$ with $\beta_n-1=O(u_n)$,

$$\forall\varepsilon>0,\quad S_n^-(\varepsilon,\beta_n)\cap S_n^+(\varepsilon,\beta_n)\cap S_n(\beta_n)\subset T_n(\varepsilon). \tag{4.4}$$

Let $G_n(\beta_n) = \{D_n^-(\beta_n,\chi)\le D_n(\chi)\le D_n^+(\beta_n,\chi), \forall\chi\in S_\Omega\}$, then (L_0) implies that $G_n(\beta_n)\subset S_n(\beta_n)$ and, from (4.4), we have

$$\forall\varepsilon>0,\quad T_n(\varepsilon)^c\subset S_n^-(\beta_n)^c\cup S_n^+(\beta_n)^c\cup G_n(\beta_n)^c$$

and hence,

$$\begin{aligned}P\left(\sup_{\chi\in S_\Omega}\left|c_{n,\chi}(D_n(\chi))-c(\chi)\right|>\varepsilon u_n\right) &\le P\left(\sup_{\chi\in S_\Omega}\left|c_{n,\chi}^-(\beta_{n,\varepsilon})-c(\chi)\right|>\varepsilon u_n\right)\\ &\quad + P\left(\sup_{\chi\in S_\Omega}\left|c_{n,\chi}^+(\beta_{n,\varepsilon})-c(\chi)\right|>\varepsilon u_n\right)\\ &\quad + P\left(1_{\{D_n^-(\beta_n,\chi)\le D_n\le D_n^+(\beta_n,\chi),\forall\chi\in S_\Omega\}}=0\right).\end{aligned}$$

This completes the proof since (L_2), (4.2) and (4.3) imply that for some $\varepsilon_0>0$

$$\sum_{n=1}^\infty P\left(\sup_{\chi\in S_\Omega}\left|c_{n,\chi}(D_n(\chi))-c(\chi)\right|>\varepsilon_0 u_n\right)<\infty.\quad\square \tag{4.5}$$

Remark 3. We wish to present two results, similar to Lemma 3, and that could be interesting for further purposes.

(i) Under the same conditions, the result stated in Lemma 3 holds by changing all the *almost complete* convergence into convergence *in probability*.
(ii) Under the same conditions, the result stated in Lemma 3 holds by changing all the $O_{\text{a.co.}}$ into $o_{\text{a.co.}}$.

The proof of (i) is the same as the one of Lemma 3, changing (4.5) into the fact that the sequence involved in (4.5) tends to zero. The proof of (ii) is similar.

4.2. Proof of Theorem 2

The main idea of the proof is to use Lemma 3, with $S_\Omega = S_{\mathcal{F}}, \mathbf{A}_i = \mathbf{X}_i,\ B_i = Y_i,\ G(t, (\chi, \mathbf{A})) = K(t^{-1}d(\chi, \mathbf{A})),\ D_n(\chi) = H_{n,k}(\chi),\ c_{n,\chi}(H_{n,k}(\chi)) = \widehat{m}_{kNN}(\chi)$ and $c(\chi) = m_L(\chi)$. We begin by recalling that the estimate

$$\widehat{m}_1(\chi, h) = \frac{1}{nEK\left(h^{-1}d(\chi, \mathbf{X}_1)\right)} \sum_{i=1}^{n} K\left(h^{-1}d(\chi, \mathbf{X}_i)\right), \quad \forall \chi \in \mathcal{F}, \tag{4.6}$$

satisfies under the conditions of Theorem 1 (see Ferraty et al. (2010))

$$\sup_{\chi \in S_{\mathcal{F}}} |\widehat{m}_1(\chi, h) - 1| = O_{\text{a.co.}}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}\left(\frac{\log n}{n}\right)}{n\phi(h)}}\right). \tag{4.7}$$

Let $\beta_n \in (0, 1)$ be an increasing sequence such that $\beta_n - 1 = O(u_n)$, with

$$u_n = \phi^{-1}(k/n)^b + \sqrt{\psi_{S_{\mathcal{F}}}(\log n/n)/k}.$$

We choose $D_n^-(\beta_n, \chi)$ and $D_n^+(\beta_n, \chi)$ such that $\varphi_\chi(D_n^-(\beta_n, \chi)) = (\sqrt{\beta_n}k)/n$, and $\varphi_\chi(D_n^+(\beta_n, \chi)) = k/(n\sqrt{\beta_n})$. Let us use the notation

$$h^-(\chi) = D_n^-(\beta_n, \chi), \qquad h^+(\chi) = D_n^+(\beta_n, \chi) \quad \text{and} \quad h = \phi^{-1}(k/n).$$

(a) *Checking* (L_4) *and* (L_5). Note that the local bandwidth $h^-(\chi)$ satisfies together with the fixed bandwidth h the conditions of Theorem 1. So (using the definition of u_n for the last equality),

$$\sup_{\chi \in S_{\mathcal{F}}} |c_{n,\chi}(D_n^-(\beta_n, \chi)) - c(\chi)| = O_{\text{a.co.}}\left(\phi^{-1}\left(\frac{k}{n}\right)^b + \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\log n/n)}{k}}\right)$$
$$= O_{\text{a.co.}}(u_n).$$

Then, (L_4) is checked. Applying the same reasoning to the case $h^+(\chi)$ we obtain

$$\sup_{\chi \in S_{\mathcal{F}}} |c_{n,\chi}(D_n^+(\beta_n, \chi)) - c(\chi)| = O_{\text{a.co.}}(u_n),$$

and (L_5) is also checked.

(b) *Checking* (L_2). Let $\chi_1, \chi_2, \ldots, \chi_{N_\varepsilon(S_{\mathcal{F}})}$ be an ε-net, with $\varepsilon = \log n/n$. For all $\eta > 0$ we can write

$$P\left(\left|1_{\{D_n^-(\beta_n,\chi) \le D_n(\chi) \le D_n^+(\beta_n,\chi), \forall \chi \in S_{\mathcal{F}}\}} - 1\right| > \eta\right)$$
$$= P\left(\inf_{\chi \in S_{\mathcal{F}}} (D_n(\chi) - D_n^-(\beta_n, \chi)) < 0\right) + P\left(\sup_{\chi \in S_{\mathcal{F}}} (D_n(\chi) - D_n^+(\beta_n, \chi)) > 0\right)$$
$$\le P\left(\min_{1 \le j \le N_\varepsilon(S_{\mathcal{F}})} (H_{n,k}(\chi_j) - D_n^-(\beta_n, \chi_j)) < 2\varepsilon\right) + P\left(\max_{1 \le j \le N_\varepsilon(S_{\mathcal{F}})} (H_{n,k}(\chi_j) - D_n^+(\beta_n, \chi_j)) > -2\varepsilon\right)$$
$$\le N_\varepsilon(S_{\mathcal{F}}) \max_{1 \le j \le N_\varepsilon(S_{\mathcal{F}})} P\left(\sum_{i=1}^{n} 1_{B(\chi_j, h^-(\chi_j)+2\varepsilon)}(\mathbf{X}_i) \ge k\right) + N_\varepsilon(S_{\mathcal{F}}) \max_{1 \le j \le N_\varepsilon(S_{\mathcal{F}})} P\left(\sum_{i=1}^{n} 1_{B(\chi_j, h^+(\chi_j)-2\varepsilon)}(\mathbf{X}_i) < k\right).$$

Then, using standard Chernoff inequalities (see, e.g., Lemma 4.3 of Burba et al. (2009)) in the right hand side of the above inequality we obtain

$$P\left(\left|1_{\{D_n^-(\beta_n) \le D_n(\chi) \le D_n^+(\beta_n), \forall \chi \in S_{\mathcal{F}}\}} - 1\right| > \eta\right) \le N_\varepsilon(S_{\mathcal{F}})\left(e^{-\log\left[C\frac{n}{k}\phi(h^-+2\varepsilon)\exp(1-\beta_n)\right]}\right)^{-k}$$
$$+ N_\varepsilon(S_{\mathcal{F}})\left(e^{\frac{\beta_n}{2}\left(1-C\frac{n}{k}\phi(h^+-2\varepsilon)\right)^2}\right)^{-k}.$$

Since $\psi_{S_{\mathcal{F}}}(\varepsilon) = \log(N_\varepsilon(S_{\mathcal{F}}))$ and $\psi_{S_{\mathcal{F}}}(\frac{\log n}{n})/k \to 0$ one has for some n_0

$$P\left(\left|1_{\{D_n^-(\beta_n)\le D_n\le D_n^+(\beta_n),\forall\chi\in S_{\mathcal{F}}\}} - 1\right| > \eta\right) \le e^{(1-\omega).\psi_{S_{\mathcal{F}}}(\log n/n)}, \quad \forall n > n_0.$$

Finally, using (H6), (L_2) is checked.

(c) *Checking* (L_3). In order to prove (L_3) we denote $f^*(\chi, h(\chi)) = EK\left(h^{-1}(\chi)d(\chi, \mathbf{X}_1)\right)$ for all $\chi \in S_{\mathcal{F}}$ and we define (omitting the argument χ)

$$F_1 = \frac{f^*\left(\chi, D_n^-(\beta_n, \chi)\right)}{f^*\left(\chi, D_n^+(\beta_n, \chi)\right)}, \qquad F_2 = \frac{\widehat{m_1}\left(\chi, h_n^-(\chi)\right)}{\widehat{m_1}\left(\chi, h_n^+(\chi)\right)} - 1, \qquad F_3 = \frac{f^*\left(\chi, h_n^+(\chi)\right)}{f^*\left(\chi, h_n^-(\chi)\right)}\beta_n - 1.$$

We have the following decomposition:

$$\left|\frac{\sum_{i=1}^n K\left((D_n^-(\beta_n,\chi))^{-1}d(\chi,\mathbf{X}_i)\right)}{\sum_{i=1}^n K\left((D_n^+(\beta_n\chi))^{-1}d(\chi,\mathbf{X}_i)\right)} - \beta_n\right| \le |F_1|\,|F_2| + |F_1|\,|F_3|.$$

Because of (H4a), one has directly that

$$\sup_{\chi\in S_{\mathcal{F}}} |F_1| < C. \tag{4.8}$$

Moreover, one can write (using (4.7) for the last equality):

$$\begin{aligned}\sup_{\chi\in S_{\mathcal{F}}} |F_2| &\le \frac{\sup_{\chi\in S_{\mathcal{F}}} |\widehat{m_1}\left(\chi, h_n^-(\chi)\right) - 1| + \sup_{\chi\in S_{\mathcal{F}}} |\widehat{m_1}\left(\chi, h_n^+(\chi)\right) - 1|}{\inf_{\chi\in S_{\mathcal{F}}} |\widehat{m_1}\left(\chi, h_n^+(\chi)\right)|},\\ &= O_{\text{a.co.}}\left(\sqrt{\frac{\psi_{S_{\mathcal{F}}}(\log n/n)}{k}}\right).\end{aligned} \tag{4.9}$$

On the other hand, following the proof of Lemma 1 in Ezzahrioui and Ould-Saïd (2008) and using (H1b) it is easily seen that for some $\tau > 0$

$$\begin{aligned}f^*(\chi, h(\chi)) &= \phi(h(\chi))\tau f(\chi) + O(\phi(h(\chi))h(\chi)^b), \quad \forall\chi \in S_{\mathcal{F}},\\ &= \tau\varphi_\chi(h(\chi)) + O(\phi(h)h^b), \quad \forall\chi\in S_{\mathcal{F}}.\end{aligned} \tag{4.10}$$

Because $\varphi_\chi(D_n^-(\beta_n,\chi))/\varphi_\chi(D_n^+(\beta_n,\chi)) = \beta_n$ one gets

$$\sup_{\chi\in S_{\mathcal{F}}} |F_3| = O(\phi(h)h^b) = O\left(\left(\sqrt{\beta_n}\phi^{-1}\left(\frac{k}{n}\right)\right)^b\right). \tag{4.11}$$

So, (L_3) is checked because $\beta_n \to 1$ and because (4.8), (4.9) and (4.11) imply that

$$\sup_{\chi\in S_{\mathcal{F}}}\left|\frac{\sum_{i=1}^n K\left((D_n^-(\beta_n,\chi))^{-1}d(\chi,\mathbf{X}_i)\right)}{\sum_{i=1}^n K\left((D_n^+(\beta_n,\chi))^{-1}d(\chi,\mathbf{X}_i)\right)} - \beta_n\right| = O_{\text{a.co.}}(u_n).$$

Note that (L_0) is obviously satisfied because of (H4a), and that (L_1) is also trivially satisfied by construction of $D_n^-(\beta_n,\chi)$ and $D_n^+(\beta_n,\chi)$. So, one can apply Lemma 3, and (4.1) is exactly the result of Theorem 2. □

5. Potential extensions

An important direct consequence of our result is the following corollary.

Corollary 4. *Under the conditions of Theorem 2, and if* $\mathbf{Z}$ *is a* $\mathcal{F}$*-valued random variable such that* $P(\mathbf{Z}\in S_{\mathcal{F}}) = 1$*, then we have*

$$|\widehat{m}_{kNN}(\mathbf{Z}) - m_L(\mathbf{Z})| = O_{\text{a.co.}}\left(\phi^{-1}\left(\frac{k}{n}\right)^b + \sqrt{\frac{\psi_{S_{\mathcal{F}}}(\log n/n)}{k}}\right).$$

In a wide range of situations, one has to develop two-stage estimation procedures. Each time that the first step of the procedure involves a functional regression estimate, one needs to use at the second stage some result like Corollary 4 (**Z** being for instance one of the previous variables $\mathbf{X}_i$ of the sample) in order to calibrate the first procedure. An example of this situation are models with both linear and nonparametric components. For those, (at least to our knowledge) the literature on *k*NN techniques is still empty, while the traditional kernel approach has been widely successful (see Aneiros Perez and Vieu (2006), Shin (2009) and Zhang and Wang (2012) for functional partial linear modelling, Ait-Sidi et al. (2008) for single functional index modelling or Ferraty et al. (2012) for functional projection pursuit regression). Corollary 4 could also be useful for data-driven smoothing parameter selection: once again this has been explored for traditional kernels (see Benheni et al. (2007)) but remains an open question for *k*NN estimates.

Since for the theoretical study of nonparametric estimates based on kernels, random ratios must be dealt with (for a deeper discussion of random ratios see Doukhan and Lang (2009)), Lemma 3 could be useful when one has to deal with random bandwidths and the desired result with fix bandwidth is known. This lemma is a generalization of a result obtained in Collomb (1980) that would allow us to obtain the rates of convergence in functional settings.

So, our conjecture is that this paper will not only be a contribution to *k*NN regression, but also a useful tool for developing other applications in functional data analysis.

Acknowledgements

Nadia Kudraszow wishes to express her deepest gratitude to Ricardo Maronna for his support. This research was partially supported by grants PIP 216 from CONICET and PICT 00899 from ANPCyT. Philippe Vieu wishes to thank all the active participants at the STAPH group in Toulouse, and more specifically Frédéric Ferraty, for long date support and comments. The authors also gratefully acknowledge the many helpful suggestions of Enrique Alvarez during the preparation of this paper and would like to thank both referees whose comments have improved significantly the presentation of this work.

References

Ait-Sidi, A., Ferraty, F., Kassa, R., Vieu, P., 2008. Cross-validated estimation in the single functional index model. Statistics 42 (6), 475–494.
Aneiros Perez, G., Vieu, P., 2006. Semi-functional partial linear model. Statist. Probab. Lett. 76 (11), 1102–1110.
Benheni, K., Ferraty, F., Rachdi, M., Vieu, P., 2007. Local smoothing regression with functional data. Comput. Statist. 22 (3), 53–71.
Bhattacharya, P., Mack, Y., 1990. Multivariate data-driven *k*-NN function estimation. J. Multivariate Anal. 35 (1), 1–11.
Biau, G., Devroye, L., 2010. On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification. J. Multivariate Anal. 101 (10), 2499–2518.
Bogachev, V.I., 1999. Gaussian Measures. In: Math Surveys and Monographs, vol. 62. Amer. Math. Soc..
Bosq, D., Lecoutre, J.P., 1987. Théorie de l'Estimation Fonctionnelle. Economica, Paris (in French).
Burba, F., Ferraty, F., Vieu, P., 2009. *k*-Nearest Neighbour method in functional nonparametric regression. J. Nonparametr. Stat. 21 (4), 453–469.
Collomb, G., 1980. Estimation de la régression par la méthode des *k* points les plus proches avec noyau: quelques propriétés de convergence ponctuelle. In: Nonparametric Asymptotic Statistics. In: Lecture Notes in Math., vol. 821. Springer, Berlin, pp. 159–175 (in French).
Devroye, L., 1982. Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. Z. Wahrscheinlichkeitstheor. Verwandte Geb. 61 (4), 467–481.
Doukhan, P., Lang, G., 2009. Evaluation for moments of a ratio with application to regression estimation. Bernoulli 15 (4), 1259–1286.
Ezzahrioui, M., Ould-Saïd, E., 2008. Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data. J. Nonparametr. Stat. 20, 3–18.
Ferraty, F., Goia, A., Salinelli, E., Vieu, P., 2012. Functional projection pursuit regression. TEST http://dx.doi.org/10.1007/s11749-012-0306-2.
Ferraty, F., Laksaci, A., Tadj, A., Vieu, P., 2010. Rate of uniform consistency for nonparametric estimates with functional variables. J. Statist. Plann. Inference 140, 335–352.
Ferraty, F., Romain, Y., 2010. The Oxford Handbook of Functional Data Analysis. Oxford University Press.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis. Theory and Practice. Springer, Berlin.
Györfi, L., Kohler, M., Krzyzak, A., Walk, H., 2002. A Distribution-Free Theory of Nonparametric Regression. In: Springer Series in Statistics, Springer-Verlag, New York.
Kolmogorov, A.N., Tikhomirov, V.M., 1959. ε-entropy and ε-capacity. Uspekhi Mat. Nauk 14, 3–86; Engl Transl. 1961. Amer. Math. Soc. Transl. Ser. 2 277–364.
Li, W.V., Shao, Q.M., 2001. Gaussian processes: inequalities, small ball probabilities and applications. In: Rao, C.R., Shanbhag, D. (Eds.), Stochastic Processes: Theory and Methods. In: Hanbook of Statistics, vol. 19. North-Holland, Amsterdam.
Lian, H., 2011. Convergence of functional *k*-nearest neighbor regression estimate with functional responses. Electron. J. Stat. 5, 31–40.
Mayer-Wolf, E., Zeitouni, O., 1993. The probability of small Gaussian ellipsoids and associated conditional moments. Ann. Probab. 21 (1), 14–24.
Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis. Springer-Verlag.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer, New York.
Shin, H., 2009. Partial functional linear regression. J. Statist. Plann. Inference 139 (10), 3405–3418.
Stone, C.J., 1982. Optimal global rates of convergence for nonparametric regression. Ann. Statist. 10 (4), 1040–1053.
van der Vaart, A.W., van Zanten, J.H., 2007. Bayesian inference with rescaled Gaussian process priors. Electron. J. Statist. 1, 433–448.
Zhang, T., Wang, Q., 2012. Semiparametric partially linear regression models for functional data. J. Statist. Plann. Inference 142 (9), 2518–2529.