Journal of Nonparametric Statistics
2012, 1–18, iFirst article

Nonparametric estimation of a surrogate density function in infinite-dimensional spaces

Frédéric Ferraty[a], Nadia Kudraszow[b]* and Philippe Vieu[a]

[a]*Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France;* [b]*Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de La Plata and CONICET, Argentina*

(*Received 11 July 2011; final version received 24 February 2012*)

A density function is generally not well defined in functional data context, but we can define a surrogate of a probability density, also called pseudo-density, when the small ball probability can be approximated by the product of two independent functions, one depending only on the centre of the ball. The aim of this paper is to study two kernel methods for estimating a surrogate probability density for functional data. We present asymptotic properties of these estimators: the convergence in probability and their rates. Simulations are given, including a functional version of smoother bootstrap selection of the parameters of the estimate.

Keywords: functional data; kernel estimators; k-nearest neighbour method; small ball probability; smoother bootstrap

AMS Subject Classifications: 62G07; 62G20

1. Introduction

Developing statistical methods for analysing functional data sets, presently called functional data analysis, has become an increasingly important area in recent years. Functional data are present in different fields such as engineering, medicine, physics, chemometrics, economy, etc. Many multivariate statistical techniques, concerning parametric models, have been extended to functional data and good overviews on this topic can be found in Ramsay and Silverman (2002, 2005) or Bosq (2000). More recently, new studies have been carried out in order to propose nonparametric methods, taking into account functional data (see Ferraty and Vieu (2006), for large discussion and references). See also Ferraty and Romain (2010) for a recent overview on functional data analysis.

In functional data analysis, the concept of probability density for a random function is an important subject of study. It is closely related to the concept of mode of the distribution of a random function, studied by Gasser, Hall, and Presnell (1998), Hall and Heckman (2002) and Ferraty and Vieu (2006, chap. 9). Delaigle and Hall (2010) proved that a probability density function generally does not exist for functional data and developed notions of density and mode

*Corresponding author. Email: nkudraszow@mate.unlp.edu.ar

ISSN 1048-5252 print/ISSN 1029-0311 online

http://dx.doi.org/10.1080/10485252.2012.671943
http://www.tandfonline.com

when functional data are considered in the space determined by the eigenfunctions of principal component analysis.

In this paper, we develop a k-nearest neighbour (kNN) method for estimating an infinite-dimensional analogue of a probability density when the small ball probability associated with the functional data can be approximated, when the radius of the ball tends to zero, as product of two independent functions, one depending only on the centre of the ball and the other on the radius.

The estimates of the surrogate density function give an useful exploratory tool for functional data set analysis and in particular for curves classification by means of estimates of the functional mode, obtained from the estimators proposed in this paper.

The principal interest in the kNN method comes from the nature of the smoothing parameter, some recent works were presented by Burba, Ferraty, and Vieu (2009) and Lian (2011). Unlike the traditional kernel method, where the smoothing parameter is a real positive number called bandwidth, the smoothing parameter in the kNN method is the number of neighbours k and takes its values in a discrete set. This feature represents a real advantage over the traditional kernel method from an implementation point of view. The other very important aspect of this method is that it allows the construction of a neighbourhood adapted to the local structure of the data.

This paper is organised as follows. In Section 2, we present our model and our surrogate for density function. In Section 3, we define a kernel estimator and prove its asymptotic properties, which will be needed in Section 4 to prove the asymptotic properties of the kNN estimator. Section 5 is devoted to some important corollaries. Computational issues including a functional bootstrapping step for automatic selection of the parameters of the estimate are discussed in Section 6. Technical proofs are reported in Appendix 1.

2. The model

Let $\{\mathbf{X}_i\}_{i=1,\dots,n}$ be n random variables independent and identically distributed as $\mathbf{X}$ and valued in $\mathcal{F}$. $(\mathcal{F}, d)$ is a semi-metric space, $\mathcal{F}$ is not necessarily of finite dimension and we do not suppose the existence of a density for the functional random variable $\mathbf{X}$, nor the existence of some dominating measure for the model.

For $\chi \in \mathcal{F}$, let $B(\chi, \varepsilon)$ be the ball of centre χ and radius ε for the topology associated with the semi-metric d:

$$B(\chi, \varepsilon) = \{\chi' \in \mathcal{F} | d(\chi', \chi) \leq \varepsilon\}.$$

We assume that there exist functions $\phi(.)$ and $f(.)$ such that for all $\chi \in \mathcal{F}$,

$$P(\mathbf{X} \in B(\chi, \varepsilon)) \sim f(\chi)\phi(\varepsilon) \quad \text{as} \ \ \varepsilon \longrightarrow 0, \tag{1}$$

which means $P(\mathbf{X} \in B(\chi, \varepsilon)) = f(\chi)\phi(\varepsilon) + o(\phi(\varepsilon))$ as $\varepsilon \to 0$. Then, the object we want to estimate is the operator f, that can be considered as an infinite-dimensional analogue of the concept of probability density, while ϕ may be interpreted as a volume parameter. In order to identify in a unique way the functions in Equation (1), we assume that

$$E(f(\mathbf{X})) = 1, \tag{2}$$

and in this way, we obtain what we shall call pseudo-density.

When $\mathcal{F} = \mathbb{R}^p$, that is, we are in the case of finite-dimensional spaces, taking f as the probability density, it can be seen that $\phi(\varepsilon) = C_p\varepsilon^p$, where C_p is the volume of the unit ball in $\mathbb{R}^p$. Some examples fulfilling the decomposition mentioned above can be found in Ferraty, Laksaci, and Vieu (2006).

Condition (1) is widely used in nonparametric functional data analysis to compensate for the lack of existence of dominating measures and of standard density function (Gasser et al. 1998; Masry 2005; Delaigle and Hall 2010; Delsol 2009).

The condition (2) is necessary for the identifiability of the model, but this is not the only possible choice. Our choice is useful to simplify the proof of the consistency of the estimators studied in this paper. Section 5 provides a direct corollary that can be applied to other identifiability conditions.

3. Kernel estimator

For a fixed $\chi \in \mathcal{F}$, the kernel estimator of $f(\chi)$ can be written as

$$\hat{f}(\chi) = \frac{(1/n)\sum_{i=1}^{n} K(h^{-1}d(\chi, \mathbf{X}_i))}{(1/n(n-1))\sum_{i=1}^{n}\sum_{j\neq i} K(h^{-1}d(\mathbf{X}_j, \mathbf{X}_i))}, \tag{3}$$

where K is an asymmetrical kernel and $h = h_n$ is a sequence of positive real numbers which decreases to zero as n goes to infinity. To establish consistency, we need some hypotheses on the distribution of $\mathbf{X}$ and on the fixed χ:

(H1) (H1a) $\forall \varepsilon > 0$, $\varphi_\chi(\varepsilon) := P(\mathbf{X} \in B(\chi, \varepsilon)) > 0$,
(H1b) $\sup_{\chi\in\mathcal{F}} |\varphi_\chi(\varepsilon)/\phi(\varepsilon) - f(\chi)| = o(1)$.

(H2) The functions $f(.)$ and $\phi(.)$ are such that
(H2a) $\phi(.)$ is increasing on a neighbourhood of zero, strictly positive and tends to zero as ε goes to zero.
(H2b) $f(.)$ is bounded and $f(\chi) > 0$.

We also need conditions on the parameters of the estimator $\hat{f}$:

(H3) The kernel $K = 1_{[0,1]}$ or K is non-negative, with compact support $[0, 1]$, the derivative K' exists on $[0, 1]$ and satisfies, for two real numbers $-\infty < C_2 < C_1 < 0$, $C_2 < K' < C_1$.
In the case of continuous kernel, we also suppose that there exists a function $\zeta_0(.)$ with $\int_0^1 \zeta_0(u)\,\mathrm{d}u > 0$ such that for all $u \in [0, 1]$,

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

(H4) The bandwidth h satisfies $\lim_{n\to\infty} 1/n\phi(h) = 0$.

Remark 3.1 (a) The existence of a function ζ_0 as in assumption (H3) is closely related to the semi-metric. The standard Ornstein–Uhlenbeck, general diffusion, general Gaussian and fractal processes are, among others, examples where such functions exist (for further details, see Ferraty, Mas, and Vieu (2007)).

(b) In the case of continuous kernel fiunctions under assumptions (H1), (H2a) and (H3), it is easily seen that

$$\exists C_3 > 0, \exists \epsilon_0, \forall \epsilon < \epsilon_0, \int_0^{\epsilon} \varphi_\chi(u)\,\mathrm{d}u > C_3\epsilon\varphi_\chi(\epsilon).$$

(c) Observe that assumptions (H1b) and (H4) imply $\lim_{n\to\infty} 1/n\varphi_\chi(h) = 0$.

Remark 3.2 The surrogate can be used to define a notion of central tendency in a population of functional variables. A functional mode χ_{mod} of f might be defined as a value $\chi \in \mathcal{F}$ which gives a local maximum of $f(.)$ (provided that exists a local maximum). In a sample S, χ_{mod} can be estimated by

$$\hat{\chi}_{\text{mod}} = \arg\max_{\boldsymbol{\xi}\in S} \hat{f}(\boldsymbol{\xi}). \tag{4}$$

This estimation is very easy to compute since in order to maximise $\hat{f}$, we just have to compute the numerator of Equation (3).

THEOREM 3.3 *Under assumptions* (H1)–(H4)*, we have*

$$\text{plim}_{n\to\infty}\hat{f}(\chi) = f(\chi),$$

where plim denotes convergence in probability.

First of all, we recall a result that will be widely used in the following. This result comes from Ferraty and Vieu (2006, p. 44, Lemmas 4.3 and 4.4) and it states that, under (H3) (in case of continuous kernel, in order that Remark 3.1 (b) holds, (H1) and (H2a) are also required), there exist constants $0 < C < C' < \infty$ such that, for h small enough

$$\forall \chi \in \mathcal{F}, \quad C\varphi_\chi(h) \leq E[K(h^{-1}d(\chi, \mathbf{X}_1))] \leq C'\varphi_\chi(h). \tag{5}$$

Proof of Theorem 3.3 Let $\hat{r}_1(\chi)$ and $\hat{r}_2(\chi)$ be the following quantities:

$$\hat{r}_1(\chi) = \frac{1}{n}\sum_{i=1}^{n} \frac{K(h^{-1}d(\chi, \mathbf{X}_i))}{f_h(\chi)},$$

$$\hat{r}_2(\chi) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} \frac{K(h^{-1}d(\mathbf{X}_j, \mathbf{X}_i))}{f_h(\chi)}, \tag{6}$$

where $f_h(\chi) = E[K(h^{-1}d(\chi, \mathbf{X}_1))]$.

Note that Equation (5) and (H1a) ensure that $f_h(\chi) > 0$. We have clearly that $\hat{f}(\chi) = \hat{r}_1(\chi)/\hat{r}_2(\chi)$. We consider the decomposition:

$$\frac{\hat{r}_1(\chi)}{\hat{r}_2(\chi)} - f(\chi) = \hat{r}_1(\chi)\left(\frac{1}{\hat{r}_2(\chi)} - f(\chi)\right) + (\hat{r}_1(\chi) - 1)f(\chi). \tag{7}$$

Using Chebyshev's inequality and Equation (5), it is easily seen that $\hat{r}_1(\chi)$ converges in probability to 1, so the second term in the right-hand side of Equation (7) converges in probability to zero. Using again that $\hat{r}_1(\chi)$ converges in probability to 1 and Lemma 3.4, we obtain the claimed result. ■

LEMMA 3.4 *Under assumptions* (H1)–(H4)

$$\text{plim}_{n\to\infty} \hat{r}_2(\chi) = \frac{1}{f(\chi)}.$$

The proof of this lemma is available in Appendix 1.

To complete the consistency property stated for the functional kernel density estimate $\hat{f}(\chi)$ defined in Equation (3), Theorem 3.5 states the rate of pointwise probability convergence. To do that, we will consider the following additional assumption:

(H1b′)] $\sup_{\chi\in\mathcal{F}} |\varphi_\chi(h)/\phi(h) - f(\chi)| = O(u_n)$.

THEOREM 3.5 *Under assumptions* (H1a), (H1b′), (H2)–(H4), *we have*

$$\hat{f}(\chi) - f(\chi) = O(u_n) + O_p\left(\frac{1}{\sqrt{n\phi(h)}}\right).$$

Remark 3.6 If the bandwidth h is chosen to satisfy $u_n \sim 1/\sqrt{n\phi(h)}$, then, Theorem 3.5 leads to

$$\hat{f}(\chi) - f(\chi) = O_p\left(\frac{1}{\sqrt{n\phi(h)}}\right).$$

Proof of Theorem 3.5 The demonstration of this result follows step by step the proof of Theorem 3.3 by using that $\hat{r}_1(\chi) - 1 = O_p(1/\sqrt{n\varphi_\chi(h)})$; the proof of this result is immediate using Chebyshev's inequality and Equation (5), together with Lemma 3.7. ∎

LEMMA 3.7 *Under assumptions* (H1a), (H1b′), (H2)–(H4), *we have*

$$\hat{r}_2(\chi) - \frac{1}{f(\chi)} = O(u_n) + O_p\left(\frac{1}{\sqrt{n}}\right).$$

The proof of this lemma can be found in Appendix 1.

It is shown that the rate of convergence is of the nonparametric type and is mainly determined by the numerator of the estimator. More precisely, Lemma 3.7 shows how the denominator has the speeder parametric rate, as it is usual with averaged nonparametric estimates (see Ferraty, Sued, and Vieu (2012), for instance).

4. *k*NN kernel estimator

For a fixed $\chi \in \mathcal{F}$, the kNN kernel estimator can be written as

$$\hat{f}_{kNN}(\chi) = \frac{(1/n)\sum_{i=1}^{n} K(H_{n,k}(\chi)^{-1} d(\chi, \mathbf{X}_i))}{(1/n(n-1))\sum_{i=1}^{n}\sum_{j\neq i} K\left(H_{n,k}(\chi)^{-1} d(\mathbf{X}_j, \mathbf{X}_i)\right)}, \tag{8}$$

where K is an asymmetrical kernel and $H_{n,k}(\chi)$ is defined as follows:

$$H_{n,k}(\chi) = \min\left\{h \in \mathbb{R}^+ : \sum_{i=1}^{n} 1_{B(\chi,h)}(\mathbf{X}_i) = k\right\}. \tag{9}$$

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

Remark 4.1 A kNN estimator of the functional mode χ_{mod} (Remark 3.2) in a sample S can be defined by replacing $\hat{f}$ with $\hat{f}_{kNN}$ in Equation (4).

To establish consistency, we need some additional assumptions:

(H2a′) $\forall\varepsilon > 0$, $\phi(\varepsilon) > 0$ with $\phi(.)$ continuous and strictly increasing on a neighbourhood of zero and $\phi(0) = 0$.
(H4′) $k = k_n$ is a sequence of positive real numbers such that $k_n \to \infty$ and $k_n/n \to 0$.

First, we state the convergence in probability of $\hat{f}_{kNN}$ defined by Equation (8).

THEOREM 4.2 *Under assumptions* (H1), (H2a′), (H2b), (H3) *and* (H4′), *we have*

$$\operatorname*{plim}_{n\to\infty} \hat{f}_{kNN}(\chi) = f(\chi).$$

Then, we establish the rate of probability convergence:

THEOREM 4.3 *Under assumptions* (H1a), (H1b′), (H2a′), (H2b), (H3) *and* (H4′), *we have*

$$\hat{f}_{kNN}(\chi) - f(\chi) = O(u_n) + O_p\left(\frac{1}{\sqrt{k}}\right).$$

Remark 4.4 If k is chosen to satisfy $u_n \sim 1/\sqrt{k}$, then, Theorem 4.3 leads to

$$\hat{f}_{kNN}(\chi) - f(\chi) = O_p\left(\frac{1}{\sqrt{k}}\right).$$

To prove Theorems 4.2 and 4.3, we need the following lemmas that give the same kind of results of Lemmas 4.1 and 4.2 of Burba et al. (2009), but using convergence in probability. For convenience, we will use the same notation as in Burba et al. (2009).

Let $(\mathbf{A}_i)_{i=1,\ldots,n}$ be n random variables valued in $(\Omega, \mathcal{A})$ where $(\Omega, \mathcal{A})$ is a general measurable space. We note $G : \mathbb{R} \times (\Omega \times \Omega) \to \mathbb{R}^+$ a measurable function such that

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

Let T be a real random variable (r.r.v.) and χ a fixed value in Ω, we define $\forall n \in \mathbb{N}\backslash\{0\}$

$$c_n(T) = \frac{\sum_{i=1}^{n} G(T, (\chi, \mathbf{A}_i))}{(1/(n-1)) \sum_{i=1}^{n} \sum_{j\neq i} G(T, (\mathbf{A}_j, \mathbf{A}_i))}.$$

Let c be a non-random positive real number, $(D_n)_{n\in\mathbb{N}}$ be a sequence of r.r.v. and $(v_n)_{n\in\mathbb{N}}$ a decreasing positive sequence.

LEMMA 4.5 (i) *If* $l = \lim v_n \neq 0$ *and if, for all increasing sequence* $\beta_n \in (0,1)$ *with* $\lim \beta_n \neq 1$, *there exist two sequences of r.r.v.* $(D_n^-(\beta_n))_{n\in\mathbb{N}}$ *and* $(D_n^+(\beta_n))_{n\in\mathbb{N}}$ *such that*
- (L_1) $D_n^-(\beta_n) \le D_n^+(\beta_n)$ $\forall n \in \mathbb{N}$ *and* $\operatorname{plim}_{n\to\infty} 1_{\{D_n^-(\beta_n)\le D_n \le D_n^+(\beta_n)\}} = 1$.
- (L_2) $\sum_{i=1}^{n} G(D_n^-(\beta_n), (\chi, \mathbf{A}_i)) / \sum_{i=1}^{n} G(D_n^+(\beta_n), (\chi, \mathbf{A}_i)) - \beta_n = O_p(v_n)$.
- (L_3) $c_n(D_n^-(\beta_n)) - c = O_p(v_n)$ *and* $c_n(D_n^+(\beta_n)) - c = O_p(v_n)$.

Then, $c_n(D_n) - c = O_p(v_n)$.

(ii) *If* $l = 0$ *and if* (L_1), (L_2) *and* (L_3) *are checked for any increasing sequence* $\beta_n \in (0,1)$ *with* $\beta_n - 1 = O(v_n)$, *then the same result holds.*

LEMMA 4.6 (i) *If* $l' = \lim v_n \neq 0$ *and if, for all increasing sequence* $\beta_n \in (0,1)$ *with* $\lim \beta_n \neq 1$, *there exist two sequences of r.r.v.* $(D_n^-(\beta_n))_{n\in\mathbb{N}}$ *and* $(D_n^+(\beta_n))_{n\in\mathbb{N}}$ *such that*
- (L_1) $D_n^-(\beta_n) \le D_n^+(\beta_n)$ $\forall n \in \mathbb{N}$ *and* $\operatorname{plim}_{n\to\infty} 1_{\{D_n^-(\beta_n)\le D_n \le D_n^+(\beta_n)\}} = 1$.
- (L_2') $\sum_{i=1}^{n} G(D_n^-(\beta_n), (\chi, \mathbf{A}_i)) / \sum_{i=1}^{n} G(D_n^+(\beta_n), (\chi, \mathbf{A}_i)) - \beta_n = o_p(v_n)$.
- (L_3') $c_n(D_n^-(\beta_n)) - c = o_p(v_n)$ *and* $c_n(D_n^+(\beta_n)) - c = o_p(v_n)$.

Then, $c_n(D_n) - c = o_p(v_n)$.

(ii) *If* $l' = 0$ *and if* (L_1), (L_2') *and* (L_3') *are checked for any increasing sequence* $\beta_n \in (0,1)$ *with* $\beta_n - 1 = O(v_n)$, *then the same result holds.*

In Appendix 1, it can be found only the sketch of proof of Lemma 4.5. The schemes of both proofs are likely the same as for Lemmas 4.1 and 4.2 of Burba et al. (2009).

Proof of Theorem 4.2 We use Lemma 4.6 (i) with $v_n = 1$, $c_n(H_{n,k}(\chi)) = \hat{f}_{kNN}(\chi)$ and $c = f(\chi)$.

Under the same conditions as in Theorem 3.3, the estimator $\hat{r}_1(\chi)$ defined in Equation (6) converges in probability to 1 and

$$\frac{1}{\phi(h)} E[K(h^{-1}d(\chi', \mathbf{X}_i))] = \alpha_1 f(\chi') + o(1) \quad \forall \chi' \in \mathcal{F}, \tag{10}$$

where α_1 is a positive suitable constant (the proof for $K = 1_{[0,1]}$ is straightforward using (H1b), and for K continuous can be found in Ezzahrioui and Ould-Saïd (2008)), combining these results, we have

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n\phi(h)} \sum_{i=1}^{n} K(h^{-1}d(\chi, \mathbf{X}_i)) = \alpha_1 f(\chi). \tag{11}$$

Let $\beta_n \in (0, 1)$, we choose $D_n^-(\beta_n)$ and $D_n^+(\beta_n)$ such that

$$\phi(D_n^-(\beta_n)) = \frac{\sqrt{\beta}}{f(\chi)} \frac{k}{n},$$
$$\phi(D_n^+(\beta_n)) = \frac{1}{\sqrt{\beta} f(\chi)} \frac{k}{n},$$

where β is the limit of the sequence β_n. Using Theorem 3.3 with $h^- = D_n^-(\beta_n) = \phi^{-1}((\sqrt{\beta}/f(\chi))(k/n))$ and $h^+ = D_n^+(\beta_n) = \phi^{-1}((1/\sqrt{\beta}f(\chi))(k/n))$, we obtain

$$c_n(D_n^-(\beta_n)) - c = o_p(1) \quad \text{and} \quad c_n(D_n^+(\beta_n)) - c = o_p(1),$$

then (L_3') is checked. Now, by applying Equation (11) both with h^- and h^+ and using probability convergence properties of ratios of random variables, we have

$$\operatorname*{plim}_{n\to\infty} \frac{\sum_{i=1}^{n} K([D_n^-(\beta_n)]^{-1} d(\chi, \mathbf{X}_i))}{\sum_{i=1}^{n} K([D_n^+(\beta_n)]^{-1} d(\chi, \mathbf{X}_i))} = \beta,$$

so (L_2') is checked. It remains just to verify (L_1). The first part is immediate. For the second one, note that for all $\varepsilon > 0$, we can write

$$P(|1_{\{D_n^-(\beta_n) \le D_n \le D_n^+(\beta_n)\}} - 1|\varepsilon) \le P\left(\sum_{i=1}^{n} 1_{B(\chi,h^-)}(\mathbf{X}_i) \ge k\right) + P\left(\sum_{i=1}^{n} 1_{B(\chi,h^+)}(\mathbf{X}_i) < k\right).$$

Then, using Lemma 4.3 of Burba et al. (2009) in the right-hand side of the above inequality, we obtain

$$P(|1_{\{D_n^-(\beta_n) \le D_n \le D_n^+(\beta_n)\}} - 1| > \varepsilon) < (\mathrm{e}^{c^- - 1 - \log(c^-)})^{-k} + (\mathrm{e}^{(1-c^+)^2/2c^+})^{-k},$$

where $c^- = k/n\varphi_x(h^-) = [\sqrt{\beta}(1 + o(1))]^{-1}$ and $c^+ = k/n\varphi_x(h^+) = \sqrt{\beta}[1 + o(1)]^{-1}$. Since $\beta \neq 1$, $c^- - 1 - \log(c^-)$ and $(1 - c^+)^2/2c^+$ are strictly positive for n large enough. From the above, it follows that

$$\lim_{n\to\infty} P(|1_{\{D_n^-(\beta_n) \le D_n \le D_n^+(\beta_n)\}} - 1| > \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

So, (L_1) is checked and the proof is complete. ■

Proof of Theorem 4.3 We can proceed analogously to the proof of Theorem 4.2, but in this case, we have to use Lemma 4.5 (ii), with $v_n = 1/\sqrt{k}$, in place of Lemma 4.6 (i).

Under the same conditions as in Theorem 3.5, the estimator $\hat{r}_1(\chi)$ defined in Equation (6) satisfies $\hat{r}_1(\chi) - 1 = O_p(1/\sqrt{n\varphi_\chi(h)})$ and

$$\frac{1}{\phi(h)} E[K(h^{-1}d(\chi', \mathbf{X}_i))] = \alpha_1 f(\chi') + O(u_n) \quad \forall \chi' \in \mathcal{F}, \tag{12}$$

where α_1 is a positive suitable constant (using (H1b′), the proof for $K = 1_{[0,1]}$ is straightforward and the proof for K continuous can be found in Ezzahrioui and Ould-Saïd (2008)), combining these results, we have

$$\frac{1}{n\phi(h)} \sum_{i=1}^{n} K(h^{-1}d(\chi, \mathbf{X}_i)) - \alpha_1 f(\chi) = O(u_n) + O_p\left(\frac{1}{\sqrt{n\phi(h)}}\right). \tag{13}$$

Let $\beta_n \in (0, 1)$ be an increasing sequence such that $\beta_n - 1 = O(1/\sqrt{k})$, we choose $D_n^-(\beta_n)$ and $D_n^+(\beta_n)$ such that

$$\phi(D_n^-(\beta_n)) = \frac{\sqrt{\beta_n}}{f(\chi)} \frac{k}{n},$$
$$\phi(D_n^+(\beta_n)) = \frac{1}{\sqrt{\beta_n} f(\chi)} \frac{k}{n}.$$

Using Theorem 3.5 with $h^- = D_n^-(\beta_n) = \phi^{-1}((\sqrt{\beta_n}/f(\chi))(k/n))$ and $h^+ = D_n^+(\beta_n) = \phi^{-1}((1/\sqrt{\beta_n} f(\chi))(k/n))$ and that β_n converges to 1, we obtain

$$c_n(D_n^-(\beta_n)) - c = O(u_n) + O_p\left(\frac{1}{\sqrt{k}}\right) \quad \text{and} \quad c_n(D_n^+(\beta_n)) - c = O(u_n) + O_p(1/\sqrt{k}),$$

then (L_3) is checked. Now, by applying Equation (13) both with h^- and h^+ and using probability convergence properties of ratios of random variables and that $\beta_n - 1 = O(1/\sqrt{k})$, we have

$$\frac{\sum_{i=1}^{n} K([D_n^-(\beta_n)]^{-1} d(\chi, \mathbf{X}_i))}{\sum_{i=1}^{n} K([D_n^+(\beta_n)]^{-1} d(\chi, \mathbf{X}_i))} - \beta_n = O(u_n) + O_p\left(\frac{1}{\sqrt{k}}\right),$$

so (L_2) is checked. The verification of (L_1) is likely the same as in the previous proof, the difference is in using that $\beta_n - 1 = O(1/\sqrt{k})$. ■

5. Some important corollaries

Corollary 5.1 provides a consistent estimator for $\phi(h)$.

COROLLARY 5.1 *Under assumptions* (H1)–(H4), *the estimator*

$$\hat{\phi}(h) = \frac{(1/n) \sum_{i=1}^{n} 1_{B(\chi,h)}(\mathbf{X}_i)}{\hat{f}(\chi)} \tag{14}$$

satisfies

$$\operatorname*{plim}_{n\to\infty} \hat{\phi}(h) - \phi(h) = 0. \tag{15}$$

And, *under assumptions* (H1a), (H1b’), (H2)–(H4), *we have*

$$\lim_{n\to\infty} \hat{\phi}(h) - \phi(h) = O(u_n \phi(h)) + O_p\left(\sqrt{\frac{\phi(h)}{n}}\right). \tag{16}$$

Proof Let $\hat{\varphi}_\chi(h) = (1/n)\sum_{i=1}^{n} 1_{B(\chi,h)}(\mathbf{X}_i)$ be the plug-in estimator of the concentration function $\varphi_\chi(h)$ and the numerator of Equation (14). We consider the decomposition

$$\hat{\phi}(h) - \phi(h) = \frac{1}{\hat{f}(\chi)}[\hat{\varphi}_\chi(h) - \varphi_\chi(h)] - \frac{1}{\hat{f}(\chi)}[\phi(h)f(\chi) - \varphi_\chi(h)] - \frac{\phi(h)}{\hat{f}(\chi)}[\hat{f}(\chi) - f(\chi)].$$

To prove Equation (15), the first term in the above decomposition will be treated by using Theorem 3.3 and that $\hat{\varphi}_\chi(h) - \varphi_\chi(h) = O_p(\sqrt{\phi(h)/n})$, the second by using (H1b) and Theorem 3.3 and the last term by using Theorem 3.3 and (H2a).

The proof of Equation (16) follows step by step the proof of Equation (15), by using Theorem 3.5 and (H1b′) in place of Theorem 3.3 and (H1b). ■

The following corollary states one of the principal advantages of the *kNN*-estimator of f when we are considering the asymmetrical box kernel.

Corollary 5.2 *If the kernel* $K = 1_{[0,1]}$, *the kNN estimator of* f *takes the simple form*

$$\hat{f}_{kNN}(\chi) = \frac{k}{\sum_{i=1}^{n} k_i/(n-1)},$$

where $k_i = \#\{j \neq i : X_j \in B(\mathbf{X}_i, H_{n,k})\}$.

As we mentioned in Section 2, the identifiability condition can be replaced by another simpler to verify. If instead the model of Section 2, we assume that there exist functions $\phi_0(.)$ and $f_0(.)$ such that for all $\chi \in \mathcal{F}$, $P(\mathbf{X} \in B(\chi, \varepsilon)) \sim f_0(\chi)\phi_0(\varepsilon)$, as $\varepsilon \to 0$, and for a fixed $\chi_0 \in \mathcal{F}$, $f_0(\chi_0) = 1$. In this model, the condition of identifiability does not depend on the distribution of $\mathbf{X}$ and $\phi_0(\varepsilon) = P(\mathbf{X} \in B(\chi_0, \varepsilon))$. The functions f and ϕ in Equation (1) differ from f_0 and ϕ_0, respectively, in a constant and therefore, the proof of Corollary 5.3 is immediate.

Corollary 5.3 *Let* $\hat{f}(\chi)$ *be the kernel estimator defined in Equation* (3) *and* $\hat{f}_{kNN}(\chi)$ *be the kNN estimator defined in Equation* (8). *Under the same assumptions of Theorems* 3.3 *and* 4.2, *respectively, the estimators* $\hat{f}_0(\chi) = \hat{f}(\chi)/\hat{f}(\chi_0)$ *and* $\hat{f}_{0,kNN}(\chi) = \hat{f}_{kNN}(\chi)/\hat{f}_{kNN}(\chi_0)$ *are consistent* (*in probability*) *for* $f_0(\chi)$. *Furthermore, under the same assumptions of Theorem* 3.5, $\hat{f}_0(\chi)$ *has the same rate of convergence of* $\hat{f}(\chi)$ *and, under the same assumptions of Theorems* 4.3, $\hat{f}_{0,kNN}(\chi)$ *has the same rate of convergence of* $\hat{f}_{kNN}(\chi)$.

6. Computational issues

From the practical point of view, the *kNN* estimator shows some advantages over the kernel estimator. One of them is that the smoothing parameter k takes its values in a discrete set, which simplifies the implementation. Moreover, the *kNN* method takes into account the local structure of the data.

Although the *k*NN estimator suffers from some standard weaknesses (the estimated pseudo-density is discontinuous and far from zero even in the case of large regions with no observed samples), it has no impact in the situation of pointwise estimation.

In this section, we propose to illustrate the behaviour of the *kNN* estimator through a simulation study. Due to the difficulty of finding an explicit expression for the small ball probability, especially for the pseudo-density, we propose only two situations. In the first situation, we consider a very simple space of smooth curves, that basically has dimension 1, in which the probability of the small ball can be computed explicitly and satisfy the model (1). In the second situation, we consider an infinite-dimensional space of rough curves (Brownian motions) but in this case, the decomposition (1) is valid only for balls centred at some smooth function χ.

In both situations, we will discuss the effects of the various parameters: kernel, number of neighbours and sample size. As usual in nonparametric statistics, the smoothing factor (here, the number of neighbours) will play a key role. A specific procedure, based on functional bootstrapping, will be introduced for selecting it in a data-driven way.

6.1. *Smooth curves*

6.1.1. *The curves and the model*

We build the simulated sample, of size 200, from the random curve constructed in the following way:

$$\mathbf{X}(t) = a\sin(t), \quad t \in [0, \pi],$$

where a is a random real variable drawn from an exponential distribution with mean 1. If we consider the standard $L^2[0,\pi]$ semi-metric

$$d(\chi, \chi') = \sqrt{\int_0^{\pi} (\chi(t) - \chi'(t))^2 \, \mathrm{d}t},$$

for an $\chi(t) = b\sin(t)$, with $b > 0$, we have

$$P(\mathbf{X} \in B(\chi, \varepsilon)) \sim \mathrm{e}^{1-b}(\mathrm{e}^{(\varepsilon\sqrt{2/\pi})-1} - \mathrm{e}^{-(\varepsilon\sqrt{2/\pi})-1}) \quad \text{as } \varepsilon \longrightarrow 0.$$

Then, the function $f_0(\chi)$ is given by $f_0(\chi) = \exp(1 - \|\chi\|/\sqrt{\pi/2})$, where $\|\cdot\|$ is the $L^2[0,\pi]$ norm, and satisfies that $f_0(b\sin(\cdot)) = \mathrm{e}^{1-b}$ for all $b > 0$. Observe that $f_0(\sin(\cdot)) = 1$.

We split the simulated sample into two sets: the learning sample $(\mathbf{X}_i)_{i=1,\ldots,100}$ used to build the estimator and the 100 fixed curves $(\mathbf{X}_i)_{i=101,\ldots,200}$ at which the true and estimated f_0 are evaluated. Moreover, the curves $\mathbf{X}_i$'s are discretized on the same grid generated from 101 equispaced measurements in $[0,\pi]$.

We predict the values $(f_0(\mathbf{X}_i))_{i=101,\ldots,200}$ using the kNN estimator $\hat{f}_{0,kNN}(\chi) = \hat{f}_{kNN}(\chi)/\hat{f}_{kNN}(\sin(\cdot))$ (see Corollary 5.3), where $\hat{f}_{kNN}(\chi)$ is defined in Equation (8). We use the standard $L^2[0,\pi]$-distance as semi-metric.

6.1.2. *Choosing the kernel K*

Figure 1 shows the predicted versus the true values of $(f_0(\mathbf{X}_i))_{i=101,\ldots,200}$, using the asymmetrical box kernel and $k = 20$ (Figure 1(a)) and the asymmetrical quadratic kernel and $k = 29$ (Figure 1(b)) in the kNN estimator. Note that MSEP is the mean square error of prediction (i.e. the sum of square errors between predicted and true values of $(f_0(\mathbf{X}_i))_{i=101,\ldots,200}$). The values of k are chosen to attain the minimum MSEP.

Since both kernels show similar results, from now on we will use the asymmetrical box kernel because the *kNN* estimator has a simple form when we use it (Corollary 5.2).

6.1.3. *Size of the sample*

Let us now look at the effect of the sample size. Since $f_0(b\sin(\cdot)) = \mathrm{e}^{1-b}$, it can be viewed as a function of b. Figure 2 shows the true function $f_0(b\sin(\cdot))$ as a function of b (solid line) and, the

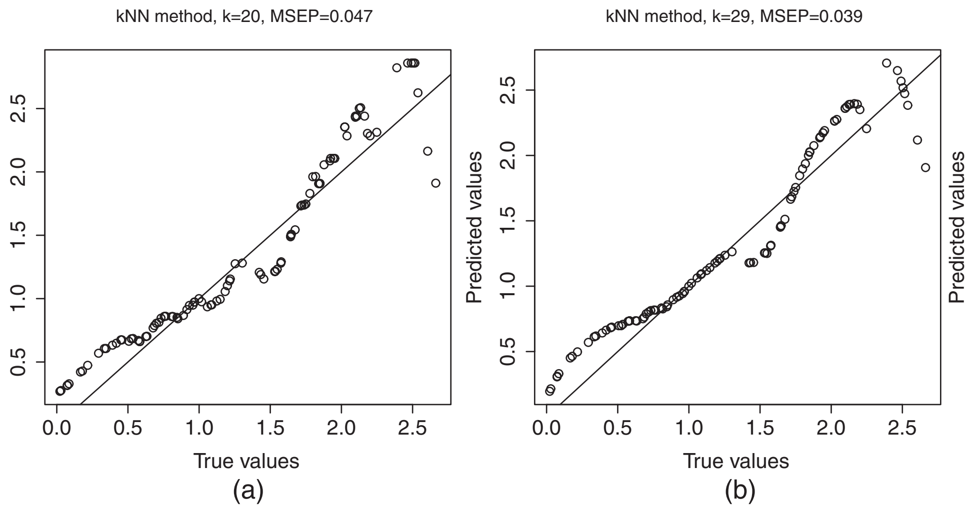

Figure 1. Predicted versus true values of $(f_0(\mathbf{X}_i))_{i=101,\ldots,200}$: (a) kNN estimator with asymmetrical box kernel and $k = 20$ and (b) kNN estimator with asymmetrical quadratic kernel and $k = 29$.

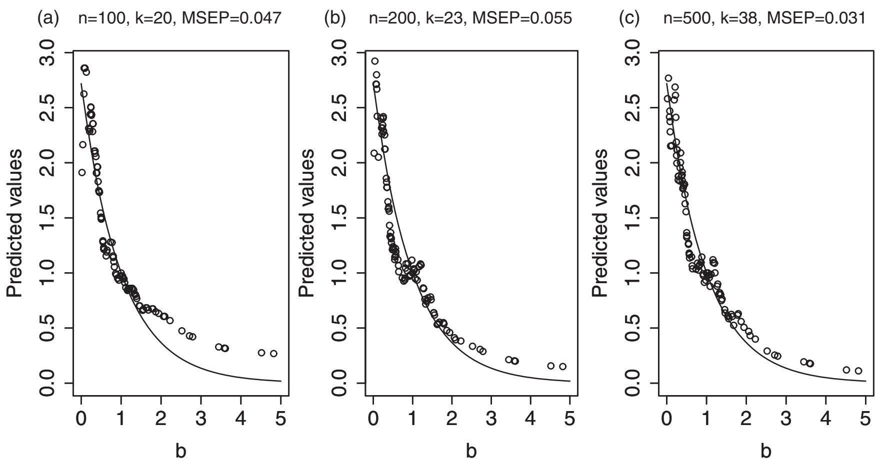

Figure 2. Predictive values of $(f_0(b_i \sin(\cdot)))_{i=101,\ldots,200}$ computed through kNN estimators. The solid line is the true function $f_0(b \sin(\cdot))$ as a function of b. Sample sizes: (a) 100, (b) 200 and (c) 500.

values b_i corresponding to the fixed curves $(\mathbf{X}_i = b_i \sin(\cdot))_{i=101,\ldots,200}$ versus the predicted values $(f_{0,kNN}(b_i \sin(\cdot)))_{i=101,\ldots,200}$ computed from three different samples of size $n = 100$ (Figure 2(a)), $n = 200$ (Figure 2(b)) and $n = 500$ (Figure 2(c)). The sample of size $n = 100$ is the learning sample $(\mathbf{X}_i)_{i=101,\ldots,200}$ used above. For each sample, the minimum MSEP is achieved for the chosen value of k. It is seen that for values of b less than or equal to 1.5, increasing the sample size does not produce an improvement, but it does for the values of b greater than 1.5. From now on, in the simulations, we will use the sample size $n = 100$.

6.1.4. *Choosing k*

We compute the values of the MSEP corresponding to the learning sample $(\mathbf{X}_i)_{i=1,\ldots,100}$ and the fixed curves $(\mathbf{X}_i = b_i \sin(\cdot))_{i=101,\ldots,200}$ for a grid of values of k. The values of the MSEP behave as a convex function between $k = 10$ and 50. The minimum MSEP is attained at $k = 20$ but between $k = 17$ and 35, the values of the MSEP are very close to the minimum. The question of data-driven choice of k is, therefore, a key one.

We propose an automatic procedure to select the number of neighbours using a kind of smoothed bootstrap and following the ideas of Faraway and Jhun (1990) for univariate data. We describe the method using the learning sample $(\mathbf{X}_i)_{i=1,\ldots,100}$ and the fixed curves $(\mathbf{X}_i)_{i=101,\ldots,200}$ used above. We construct initial kNN estimates of the true values $(f_0(\mathbf{X}_i))_{i=101,\ldots,200}$ with k_0 as the number of neighbours, they will be denoted by $(\hat{f}_{0,kNN}(\mathbf{X}_i; k_0))_{i=101,\ldots,200}$. For $i = 101, \ldots, 200$, we have $h_i := H_{n,k_0}(\mathbf{X}_i)$, with $H_{n,k}(\chi)$ defined in Equation (9). The next step is based on a kind of smoothed bootstrap; for each discretised curve in the learning sample $(\mathbf{X}_i(t_1), \ldots, \mathbf{X}_i(t_{101}))$, we defined $\mathbf{X}^i_{m,r}(t_j) = \mathbf{X}^*_{m,r}(t_j) + \mathbf{Z}_i(t_j)$, where $(\mathbf{X}^*_{m,r}(t_1), \ldots, \mathbf{X}^*_{m,r}(t_{101}))_{m=1,\ldots,100}$ is a standard bootstrap replication drawn from the original discretised trajectories of the learning sample for $r = 1, \ldots, B$, where B is the number of bootstrap samples to be taken, and $(\mathbf{Z}_i(t_1), \ldots, \mathbf{Z}_i(t_{101}))$ is normally distributed with the mean zero and covariance matrix $h_i \boldsymbol{\Sigma}_x$, where $\boldsymbol{\Sigma}_x$ is the sample covariance of the discretised learning sample (this construction is adapted from Cuevas, Febrero, and Fraiman (2006)). We obtain the bootstrap choice of the number of neighbours $\hat{k}_B$ by minimising the expression

$$\text{BMSE}(k) = (100B)^{-1} \sum_{i=1}^{100} \sum_{r=1}^{B} (\hat{f}^*_{0,kNN}(\mathbf{X}_i; k) - \hat{f}_{0,kNN}(\mathbf{X}_i; k_0))^2,$$

where $\hat{f}^*_{0,kNN}(\mathbf{X}_i; k)$ is the kNN estimator of $(f_0(\mathbf{X}_i))$ computed using the smoothed bootstrap sample $(\mathbf{X}^i_{m,r}(t_1), \ldots, \mathbf{X}^i_{m,r}(t_{101}))_{m=1,\ldots,100}$ and with k as the number of neighbours.

To improve the computational time of the method, we can use only some of the fixed curves. They must be chosen carefully, in order to obtain a representative subset (for example, we can reorder the fixed curves with respect to their seminorms and conserve a subset of equispaced curves). For the simulated sample, we used $k_0 = 50$ (half of the learning sample size) and $B = 10$ and we obtained $\hat{k}_B = 18$ using only 21 fixed curves and $\hat{k}_B = 25$ using the complete set of fixed curves, both bootstrap choices give reasonable MSEPs.

It is possible to iterate this bootstrap method, that is, use the bootstrap choice of the number of neighbours as the new initial choice and apply the method again. Table 1 shows the number of neighbours $\hat{k}_B$ corresponding to the first (second row) and second iteration (fifth row) of the bootstrap method for 10 different samples simulated as we described at the beginning of this section (a learning sample with sample size $n = 100$ and a sample of fixed curves with the same sample size). The bootstrap method was applied with the same parameters as above and using

Table 1. Number of neighbours $\hat{k}_B$ corresponding to the first (second row) and second iteration (fifth row) of the bootstrap method for 10 different simulated samples.

Sample	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
First iteration										
k_{opt}	32	32	10	30	13	20	6	29	16	24
$\hat{k}_B$	19	18	13	16	19	14	17	17	13	17
MSEP	0.05	0.36	0.28	0.21	0.14	0.22	0.13	0.25	0.15	0.08
Bootstrap MSEP	0.07	0.36	0.31	0.27	0.22	0.5	0.31	0.39	0.20	0.09
Second iteration										
$\hat{k}_B$	16	17	12	16	18	16	16	17	13	19
Bootstrap MSEP	0.13	0.30	0.28	0.27	0.22	0.28	0.28	0.40	0.20	0.09

Notes: The first row shows the number of neighbours that minimises the MSEP, the fourth and sixth rows correspond to the MSEP obtained from the first and second iteration of the bootstrap method, respectively, and the third row shows the minimum MSEPs.

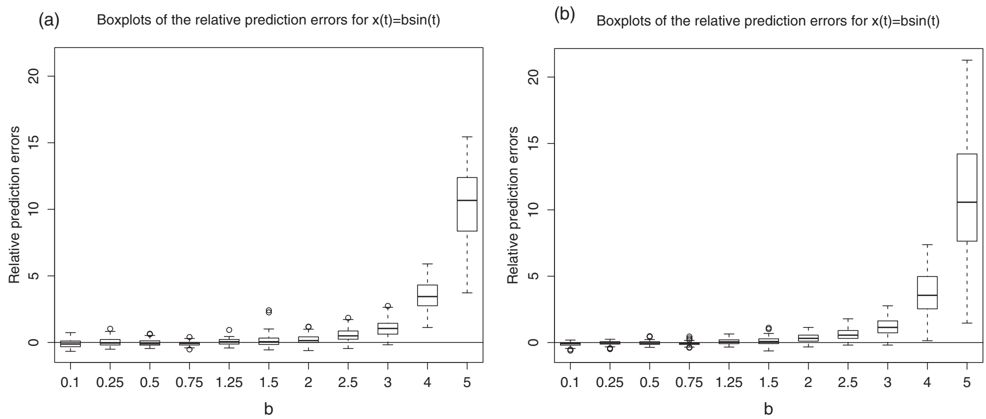

Figure 3. Boxplots of the relative errors over curves of the form $\chi_b(t) = b\sin(t)$, each boxplot corresponds to a value of b specified in the x-axis. (a) Results corresponding to the bootstrap method and (b) the number of neighbours minimising the MSEP.

only 21 fixed curves. Also are shown the number of neighbours that minimises the MSEP (first row) and the MSEP obtained for each choice of k (fourth and sixth rows correspond to the MSEP obtained from the first and second iteration of the bootstrap method, respectively, and the third row shows the minimum MSEPs). For the 10 simulated samples, the plots of the MSEP versus the number of neighbours k have a range of values of k containing the optimal choice where the variation of the MSEP is convex and very small. All the bootstrap choices lie within this range. It is seen that the second iteration does not produce an improvement with respect to the first iteration, this supports the fact that the bootstrap method does not depend significantly on the initial value of k_0.

6.1.5. *Final analysis*

Finally, we simulate 50 samples of curves as described above with sample sizes $n = 100$, $\{(X_{ij})_{j=1,\dots,101}\}_{i=1,\dots,50}$. For each sample we compute, for a set of fixed curves of the form $\chi_b(t) = b\sin(t)$, the relative errors $(\hat{f}_0(\chi_b) - f_0(\chi_b))/f_0(\chi_b)$. The kernel used to compute $\hat{f}_0(.)$ is the asymmetrical box kernel and for each sample, the number of neighbours k is chosen through the bootstrap method described above using $B = 10$. According to the analysis above, the bootstrap method is stable with respect to the choice of the initial parameter, so we use $k_0 = 50$ and we only perform one iteration of the method.

The results are reported by means of boxplots in Figure 3. Figure 3(a) shows the results for the bootstrap method and Figure 3(b) the results corresponding to the number of neighbours that minimises the MSEP. In both panels, the relative errors are small for values of b less than or equal to 1.5, and increase with b for $b > 1.5$. This is because the probability of occurrence decreases as b increases.

Table 2 shows the MSEP for each curve of the form $\chi_b(t) = b\sin(t)$, over the 50 simulated samples, using the bootstrap method (first row) and minimising the MSEP (second row).

6.2. *Brownian motion*

Let $\mathbf{X} = \{X_t\}_{t\in[0,1]}$ be the standard Brownian motion. The standard Brownian motion can be viewed as a map into the space $L^2[0,1]$. In this case, Theorem 3.1 of Li and Shao (2001) provides a similar result for small ball probabilities. For all absolutely continuous function $\chi : [0,1] \to \mathbb{R}$

Table 2. MSEP for each curve $\chi_b(t) = b\sin(t)$ over the 50 simulated samples.

b	0.1	1.25	0.5	0.75	1.25	1.5	2	2.5	3	4	5
BMSEP(b)	0.69	0.49	0.18	0.06	0.04	0.12	0.03	0.03	0.03	0.04	0.05
MSEP(b)	0.23	0.13	0.09	0.05	0.03	0.05	0.03	0.03	0.03	0.03	0.04

Note: The first row corresponds to the results from bootstrap method and the second to the number of neighbours minimising the MSEP.

with $\chi(0) = 0$ and $\int_0^1 \chi'(t)^2\,\mathrm{d}t < \infty$, the small ball probabilities of Brownian motion satisfy

$$P(\mathbf{X} \in B(\chi, \varepsilon)) \sim \exp\left\{-\frac{1}{2}\int_0^1 \chi'(t)^2\,\mathrm{d}t\right\} P(\mathbf{X} \in B(0, \varepsilon)) \quad \text{as } \varepsilon \longrightarrow 0.$$

Note that the above equation does not correspond exactly to the model (1) because the centre of the ball is not a Brownian motion. However, it can be of interest to estimate the function $f_0(\chi)$ given by

$$f_0(\chi) = \exp\left\{-\frac{1}{2}\int_0^1 \chi'(t)^2\,\mathrm{d}t\right\},$$

where χ is an absolutely continuous function.

We will use the proposed kNN estimator to estimate from a Brownian motion sample the values of the function f_0 (which satisfies $f_0(0) = 1$) evaluated in a sample of absolutely continuous functions (fixed curves). The results will be compared with the true values of f_0 in the sample of absolutely continuous curves. In order to do that, we consider two simulated standard Brownian motion samples $\{\mathbf{X}_i\}_{i=1}^n$, one of size $n = 100$ and the other of size $n = 1000$. The curves $\mathbf{X}_i$'s are rough and are discretised on the same grid generated from 100 equispaced measurements in [0, 1].

We also consider a simulated sample of $m = 100$ random curves of the form $\mathbf{Y}_i(t) = \int_0^t W_i(s)\,\mathrm{d}s$, where $W_i(s)$ is a Brownian motion. The curves of the sample $(\mathbf{Y}_i(t))_{i=1,\ldots,100}$ are absolutely continuous with $\mathbf{Y}_i(0) = 0$ and $\int_0^1 \mathbf{Y}_i'(t)^2\,\mathrm{d}t < \infty$ so, we can estimate the values of $(f_0(\mathbf{Y}_i(t)))_{i=1,\ldots,100}$. The curves $\mathbf{Y}_i$'s are discretised on the same grid of the $\mathbf{X}_i$'s. We are obliged to consider a sample from a different population process in order to compare the estimates obtained in the simulation with their true values. But in a practical context, we can estimate the surrogate density corresponding to model (1) in any standard Brownian motion path.

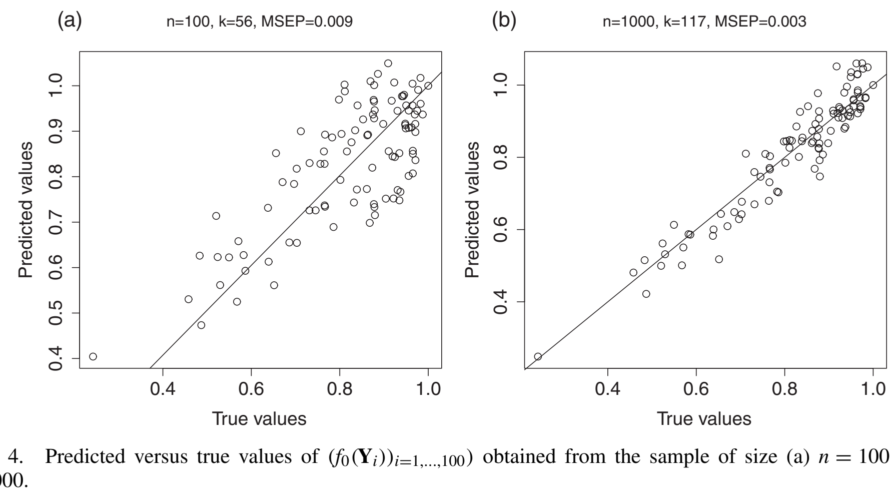

Figure 4. Predicted versus true values of $(f_0(\mathbf{Y}_i))_{i=1,\ldots,100}$ obtained from the sample of size (a) $n = 100$ and (b) $n = 1000$.

We predict the values $(f_0(\mathbf{Y}_i))_{i=1,\dots,100}$ using the kNN kernel estimator $\hat{f}_{0,kNN}(\chi) = \hat{f}_{kNN}(\chi)/\hat{f}_{kNN}(0)$ (Corollary 5.3), where $\hat{f}_{kNN}(\chi)$ is defined in Equation (8), with each sample $\{\mathbf{X}_i\}_{i=1}^n$. We use the standard $L^2[0,1]$-distance as semi-metric and the asymmetrical box kernel. The number of neighbours was chosen to achieve the minimum MSEP. The results are shown in Figure 4.

In this case, the results obtained from the sample with $n = 1000$ are much better than the results obtained from the sample with $n = 100$. The MSEP and the ratio between the optimal number of neighbours and n corresponding to $n = 1000$ are significantly smaller than those corresponding to $n = 100$. Note that, although the model (1) is not satisfied exactly, the performance of the kNN estimator is quite good, especially for large n.

7. Conclusions

In this paper, we have presented kernel and kNN kernel estimators of a surrogate for density of a random function. We show the pointwise convergence in probability of these estimators and we established their rates of convergence. It is important to note that both rates are similar, but the kNN method is more natural in the infinite-dimension context. Its main qualities are that the smoothing parameter k takes its values in a discrete set, it has a simple form when the asymmetrical box kernel is used and it takes into account the local structure of the data. Simulations show that the proposed kNN estimator combines both easiness of implementation and good behaviour on finite sample sizes, as well for smooth curves as for rough ones. Of course, treatment of rough curves needs higher sample size to deal with the computational issues, a functional bootstrapping method is proposed.

Acknowledgements

Frédéric Ferraty and Philippe Vieu wish to thank all the active participants at the STAPH group in Toulouse for their long date support and pertinent comments. 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.

References

Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, Lecture Notes in Statistics, Vol. 149, New York: Springer.

Burba, F., Ferraty, F., and Vieu, P. (2009), 'k-Nearest Neighbour Method in Functional Nonparametric Regression', *Journal of Nonparametric Statistics*, 21, 453–469.

Cuevas, A., Febrero, M., and Fraiman, R. (2006), 'On the Use of the Bootstrap for Estimating Functions with Functional Data', *Computational Statistics & Data Analysis*, 51, 1063–1074.

Delaigle, A., and Hall, P. (2010), 'Defining Probability Density for a Distribution of Random Functions', *The Annals of Statistics*, 38(2), 1171–1193.

Delsol, L. (2009), 'Advances on Asymptotic Normality in Non-Parametric Functional Time Series Analysis', *Statistics*, 43(1), 13–33.

Ezzahrioui, M., and Ould-Saïd, E. (2008), 'Asymptotic Normality of a Nonparametric Estimator of the Conditional Mode Function for Functional Data', *Journal of Nonparametric Statistics*, 20, 3–18.

Faraway, J., and Jhun, M. (1990), 'Bootstrap Choice of Bandwidth for Density Estimation', *Journal of the American Statistical Association*, 85, 1119–1122.

Ferraty, F., and Romain, Y. (2010), *The Oxford Handbook of Functional Data Analysis*, Oxford: Oxford University Press.

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis. Theory and Practice*, Berlin: Springer.

Ferraty, F., Laksaci, A., and Vieu, P. (2006), 'Estimating Some Characteristics of the Conditional Distribution in Nonparametric Functional Models', *Statistical Inference for Stochastic Processes*, 9, 47–76.

Ferraty, F., Mas, A., and Vieu, P. (2007), 'Nonparametric Regression on Functional Data: Inference and Practical Aspects', *Australian & New Zealand Journal of Statistics*, 49(3), 267–286.

Ferraty, F., Sued, M., and Vieu, P. (2012), 'Mean estimation with data missing at random for functional covariables', *Statistics*, http://dx.doi.org/10.1080/02331888.2011.650172.

Gasser, T., Hall, P., and Presnell, B. (1998), 'Nonparametric Estimation of the Mode of a Distribution of Random Curves', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 681–691.
Hall, P., and Heckman, N. (2002), 'Estimating and Depicting the Structure of a Distribution of Random Functions', *Biometrika*, 89, 145–158.
Lee, A.J. (1990), *U-Statistics: Theory and Practice*, New York: Marcel Dekker Inc.
Li, W.V., and Shao, Q.M. (2001), 'Gaussian Processes: Inequalities, Small Ball Probabilities and Applications', in *Stochastic Processes: Theory and Methods. Handbook of Statistics* (Vol. 19), eds. C.R. Rao and D. Shanbhag, Amsterdam: North-Holland, pp. 533–598.
Lian, H. (2011), 'Convergence of Functional *k*-Nearest Neighbor Regression Estimate with Functional Responses', *Electronic Journal of Statistics*, 5, 31–40.
Masry, E. (2005), 'Nonparametric Regression Estimation for Dependent Functional Data: Asymptotic Normality', *Stochastic Processes and their Applications*, 115(1), 155–177.
Ramsay, J.O., and Silverman, B.W. (2002), *Applied Functional Data Analysis*, Berlin: Springer.
Ramsay, J.O., and Silverman, B.W. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer.

Appendix 1. Proofs of technical lemmas

A.1. *Proof of Lemma* 3.4

To get the claimed result, the main probabilistic tool will be the use of general results for U-statistics. So, in order to follow the traditional notation in this area, we write

$$\hat{r}_2(\chi) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{i<j} \frac{K(h^{-1} d(\mathbf{X}_j, \mathbf{X}_i))}{f_h(\chi)}$$

and then we can see that $\hat{r}_2(\chi)$ is an U-statistic of order 2 with kernel $F_{\chi,h}(\mathbf{X}_j, \mathbf{X}_i) := K(h^{-1} d(\mathbf{X}_j, \mathbf{X}_i))/f_h(\chi)$.

Let us consider

$$\theta_h = E[\hat{r}_2(\chi)] = E\left[\frac{K(h^{-1} d(\mathbf{X}_2, \mathbf{X}_1))}{f_h(\chi)}\right] = \frac{E[f_h(\mathbf{X}_2)]}{f_h(\chi)}$$

and $r^*(\chi) = (2/n) \sum_{i=1}^{n} (f_h(\mathbf{X}_i)/f_h(\chi) - \theta_h)$.

Writing

$$\hat{r}_2(\chi) - \frac{1}{f(\chi)} = (\hat{r}_2(\chi) - \theta_h - r^*(\chi)) + r^*(\chi) + \theta_h - \frac{1}{f(\chi)}, \tag{A1}$$

we obtain that Lemma 3.4 is a consequence of the following intermediate results:

$$\hat{r}_2(\chi) - \theta_h - r^*(\chi) = O_p\left(\frac{1}{n\sqrt{\phi(h)}}\right), \tag{A2}$$

$$r^*(\chi) = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{A3}$$

and

$$\theta_h - \frac{1}{f(\chi)} = o(1). \tag{A4}$$

The assertion in Equation (A2) can be deduced from

$$\mathrm{Var}(\hat{r}_2(\chi) - \theta_h - r^*(\chi)) = O\left(\frac{1}{n^2 \phi(h)}\right). \tag{A5}$$

This is proved by writing

$$\begin{aligned} \mathrm{Var}(\hat{r}_2(\chi) - \theta_h - r^*(\chi)) &= \mathrm{Var}(r^*(\chi) - (\hat{r}_2(\chi) - \theta_h)) \\ &= \mathrm{Var}(r^*(\chi)) - 2\,\mathrm{Cov}(r^*(\chi), \hat{r}_2(\chi)) + \mathrm{Var}(\hat{r}_2(\chi)). \end{aligned} \tag{A6}$$

Let $\sigma_1^2(\chi, h) = \mathrm{Var}(f_h(\mathbf{X}_2)/f_h(\chi))$ and $\sigma_2^2(\chi, h) = \mathrm{Var}(F_{\chi,h}(\mathbf{X}_2, \mathbf{X}_1))$. Independence between observations guarantees that

$$\mathrm{Var}(r^*(\chi)) = \frac{2^2}{n} \sigma_1^2(\chi, h). \tag{A7}$$

On the other hand, the well-known formula for the variance of U-statistics (Lee 1990, Theorem 3, sec. 1.3), implies that

$$\mathrm{Var}(\hat{r}_2(\chi)) = 2^2\frac{n-2}{n(n-1)}\sigma_1^2(\chi,h) + \frac{2}{n(n-1)}\sigma_2^2(\chi,h). \tag{A8}$$

Finally, since $\mathrm{Cov}(f_h(\mathbf{X}_i), f_h(\mathbf{X}_j)) = f_h(\chi)^2\sigma_1^2(\chi,h)\delta_{ij}$ and for $k \neq l$

$$\mathrm{Cov}(f_h(\mathbf{X}_i), K(h^{-1}d(\mathbf{X}_h,\mathbf{X}_l))) = \mathrm{Cov}(f_h(\mathbf{X}_i), f_h(\mathbf{X}_k))\delta_{ik} + \mathrm{Cov}(f_h(\mathbf{X}_i), f_h(\mathbf{X}_l))\delta_{il},$$

where δ_{mr} is the Kronecker delta, we get

$$\mathrm{Cov}(r^*(\chi), \hat{r}_2(\chi)) = \frac{2^2}{n}\sigma_1^2(\chi,h). \tag{A9}$$

Substituting Equations (A7), (A8) and (A9) into Equation (A6), we obtain

$$\mathrm{Var}(\hat{r}_2(\chi) - \theta_h - r^*(\chi)) \leq \frac{2\sigma_2^2(\chi,h)}{n(n-1)} \leq \frac{2}{n(n-1)}\frac{E[K^2(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h^2(\chi)}. \tag{A10}$$

If $K = 1_{[0,1]}$, by (H1b) and (H2b), we have

$$\frac{E[K^2(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h^2(\chi)} = \frac{E[\varphi_{\mathbf{X}_2}(h)]}{\varphi_\chi^2(h)} = \frac{E[f(\mathbf{X}_2)] + o(1)}{\phi(h)[f(\chi)+o(1)]^2} = O\left(\frac{1}{\phi(h)}\right). \tag{A11}$$

On the other hand, if K is a continuous kernel, following the proof of Lemma 1 in Ezzahrioui and Ould-Saïd (2008) and using (H1b), it is easily seen that

$$\sup_{\chi\in\mathcal{F}}\left|\frac{1}{\phi(h)}E[K(h^{-1}d(\chi,\mathbf{X}_i))] - \alpha_1 f(\chi)\right| = o(1), \tag{A12}$$

where α_1 is the same as the constant in Equation (10). Then, using Equation (A12) and (H2b), we get

$$\frac{E[K^2(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h^2(\chi)} \leq \left(\max_{u\in[0,1]} K(u)\right)\frac{\alpha_1 E[f(\mathbf{X}_2)] + o(1)}{\phi(h)[\alpha_1 f(\chi)+o(1)]^2} = O\left(\frac{1}{\phi(h)}\right). \tag{A13}$$

From Equations (A10), (A11) and (A13), we deduce Equation (A5).

The proof of Equation (A3) follows from the fact that under (H1), (H2b) and (H3)

$$\frac{f_h(\mathbf{X}_i)}{f_h(\chi)} \leq \frac{C'\varphi_{\mathbf{X}_i}(h)}{C\varphi_\chi(h)} \leq \frac{C'(f(\mathbf{X}_i)+o(1))}{C(f(\chi)-o(1))} \leq \frac{C'(\sup f + o(1))}{C(f(\chi)-o(1))},$$

and so $f_h(\mathbf{X}_i)/f_h(\chi)$ is a bounded random variable. It remains just to prove Equation (A4). For that, first consider $K = 1_{[0,1]}$; by (H1b) and the fact that $E[f(\mathbf{X}_2)] = 1$, we have

$$\theta_h = \frac{E[f_h(\mathbf{X}_2)]}{f_h(\chi)} = \frac{E[\varphi_{\mathbf{X}_2}(h)]}{\varphi_\chi(h)} = \frac{E[f(\mathbf{X}_2)] + o(1)}{f(\chi)+o(1)} \longrightarrow \frac{1}{f(\chi)}, \tag{A14}$$

as $n \to \infty$. On the other hand, if K is a continuous kernel that satisfies (H3), using Equation (A12) and $E[f(\mathbf{X}_2)] = 1$, we get

$$\theta_h = \frac{E[K(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h(\chi)} = \frac{\alpha_1 E[f(\mathbf{X}_2)] + o(1)}{\alpha_1 f(\chi)+o(1)} \longrightarrow \frac{1}{f(\chi)}, \tag{A15}$$

as $n \to \infty$. Then, Equations (A14) and (A15) prove the desired result.

A.2. *Proof of Lemma* 3.7

Following the beginning of the proof of Lemma 3.4, we have again Equation (A1) and then we obtain that Lemma 3.7 is a consequence of the intermediate results (A2), (A3) and

$$\theta_h - \frac{1}{f(\chi)} = O(u_n). \tag{A16}$$

As in Lemma 3.4, to prove assertion (A2) is sufficient to show that

$$\frac{E[K^2(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h^2(\chi)} = O\left(\frac{1}{\phi(h)}\right). \tag{A17}$$

If $K = 1_{[0,1]}$, by (H1b′) and (H2b), analysis similar to that in Equation (A11), implies Equation (A17). On the other hand, if K is a continuous kernel, following the proof of Lemma 1 in Ezzahrioui and Ould-Saïd (2008) and using (H1b), it is easily seen that

$$\sup_{\chi\in\mathcal{F}}\left|\frac{1}{\phi(h)}E[K(h^{-1}d(\chi,\mathbf{X}_i))]-\alpha_1 f(\chi)\right|=O(u_n),\tag{A18}$$

where α_1 is the same as the constant in Equation (12). Then, using Equation (A18) and (H2b), we get

$$\frac{E[K^2(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h^2(\chi)}\leq\left(\max_{u\in[0,1]}K(u)\right)\frac{\alpha_1 E[f(\mathbf{X}_2)]+O(u_n)}{\phi(h)[\alpha_1 f(\chi)+O(u_n)]^2}=O\left(\frac{1}{\phi(h)}\right).$$

The proof of Equation (A3) follows from the fact that under (H1a), (H1b′), (H2b) and (H3)

$$\frac{f_h(\mathbf{X}_i)}{f_h(\chi)}\leq C\frac{\varphi_{\mathbf{X}_i}(h)}{\varphi_\chi(h)}\leq C\frac{f(\mathbf{X}_i)+O(u_n)}{f(\chi)-O(u_n)}=O(1).$$

It remains just to prove Equation (A16). For that, first consider $K=1_{[0,1]}$; by (H1b′) and the fact that $E[f(\mathbf{X}_2)]=1$, we have

$$\theta_h=\frac{E[f_h(\mathbf{X}_2)]}{f_h(\chi)}=\frac{E[\varphi_{\mathbf{X}_2}(h)]}{\varphi_\chi(h)}=\frac{E[f(\mathbf{X}_2)]+O(u_n)}{f(\chi)+O(u_n)}=\frac{1+O(u_n)}{f(\chi)+O(u_n)}=\frac{1}{f(\chi)}+O(u_n).\tag{A19}$$

On the other hand, if K is a continuous kernel that satisfies (H3), using Equation (A18) and $E[f(\mathbf{X}_2)]=1$, we get

$$\theta_h=\frac{E[K(h^{-1}d(\mathbf{X}_2,\mathbf{X}_1))]}{f_h(\chi)}=\frac{\alpha_1 E[f(\mathbf{X}_2)]+O(u_n)}{\alpha_1 f(\chi)+O(u_n)}=\frac{\alpha_1+O(u_n)}{\alpha_1 f(\chi)+O(u_n)}=\frac{1}{f(\chi)}+O(u_n).\tag{A20}$$

Then, Equations (A19) and (A20) prove the desired result.

A.3. *Sketch of proof Lemma* 4.5

For all sequences $\beta_n\in(0,1)$, (L_2) and (L_3) give

$$c_n^-(\beta_n)=\frac{\sum_{i=1}^n G(D_n^-(\beta_n),(\chi,\mathbf{A}_i))}{(1/(n-1))\sum_{i=1}^n\sum_{j\neq i}G(D_n^+(\beta_n),(\mathbf{A}_j,\mathbf{A}_i))}=c\beta_n+O_p(v_n)\tag{A21}$$

and

$$c_n^+(\beta_n)=\frac{\sum_{i=1}^n G(D_n^+(\beta_n),(\chi,\mathbf{A}_i))}{(1/(n-1))\sum_{i=1}^n\sum_{j\neq i}G(D_n^-(\beta_n),(\mathbf{A}_j,\mathbf{A}_i))}=\frac{c}{\beta_n}+O_p(v_n).\tag{A22}$$

Under (ii), we chose

$$\beta_n=\beta_{n,m}=1-\frac{v_n}{3cm}\quad\forall m>m_0=\frac{v_1}{3c},\tag{A23}$$

whereas, under (i), we take

$$\beta_n=\beta_{n,m}=1-\frac{l}{3cm}\quad\forall m>m_0=\frac{l}{3c}.\tag{A24}$$

For all $m>0$, we define $T_n(m)$, $S_n^-(m,\beta_n)$, $S_n^+(m,\beta_n)$, $S_n(\beta_n)$, $G_n^-(m)$, $G_n^+(m)$ and $G_n(m)$ as in the proof of Lemma 4.1 of Burba *et al.* (2009) by replacing ϵ by m. Then, using (L_0) and Equations (A23) and (A24), we have that

$$\forall m>m_0,\quad T_n(m)^c\subset G_n^-(m)^c\cup G_n^+(m)^c\cup G_n(m)^c$$

and hence

$$P(|c_n(D_n)-c|>mv_n)\leq P\left(|c_n^-(\beta_{n,m})-\beta_{n,m}c|>\frac{mv_n}{3}\right)+P\left(\left|c_n^+(\beta_{n,m})-\frac{c}{\beta_{n,m}}\right|>\frac{mv_n}{3}\right)$$
$$+P(D_n\in[D_n^-(\beta_{n,m}),D_n^+(\beta_{n,m})]).$$

By (L_0) and Equations (A21) and (A22), we have $\lim_m\underline{\lim}_{n\to\infty}P(|c_n(D_n)-c|>mv_n)=0$, and the lemma follows.