Journal of Multivariate Analysis 102 (2011) 73–86

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

On local times, density estimation and supervised classification from functional data

P. Llop [a,*,1], L. Forzani [b], R. Fraiman [c,d]

[a] *Instituto de Matemática Aplicada del Litoral, Argentina*

[b] *Instituto de Matemática Aplicada del Litoral - CONICET - Universidad Nacional del Litoral, Argentina*

[c] *Departamento de Matemática, Universidad de San Andrés, Argentina*

[d] *CMAT, Universidad de la República, Uruguay*

ARTICLE INFO

Article history:
Received 18 December 2009
Available online 13 August 2010

AMS subject classifications:
62G07
62H30
62G20

Keywords:
Functional data
Density estimation
Nearest neighbors

ABSTRACT

In this paper, we define a $\sqrt{n}$-consistent nonparametric estimator for the marginal density function of an order one stationary process built up from a sample of i.i.d continuous time trajectories. Under mild conditions we obtain strong consistency, strong orders of convergence and derive the asymptotic distribution of the estimator. We extend some of the results to the non-stationary case. We propose a nonparametric classification rule based on local times (occupation measure) and include some simulations studies.

1. Introduction

In the past few years, advances in technology and modern computing environments allowed us to collect and analyze high-dimensional data coming from different fields such as health sciences, engineering, physical sciences, genetics, geophysics, chemometrics, finance and social sciences. When the data are recorded densely over a period of time, often by machine, they are thought of as curves and, in this context, each observed curve corresponds to a different individual in the sample. Since the data come in a discretized way, we could treat them as vectors in $\mathbb{R}^d$; however, we choose to treat them as curves in order to gain more information and, in that way, obtain better estimators. The set of tools used for analyzing this kind of data is called functional data analysis (or FDA, an acronym coined by Ramsay and Dalzell [34]).

FDA is a very popular topic in the modern statistics (see [15]) and a great effort has been (and still is) made to provide statistical tools for its analysis. This popularity is due to its elegant theory and its practical applications (see [20]). For instance, the recent books by Ramsay and Silverman [35,36] and Ferraty and Vieu [17], present important theoretical results as well as efficient statistical packages which make them the most useful monographs existing in this area.

The most common approaches to treat the discretized data in functional data are those that start by a *regularization procedure* [21] or a *filtering method* (which leads to replacement of every function by its coefficients with respect to the basis of a suitable finite-dimensional subspace) and then continue with the theoretical analysis using those smoothed curves. The principal feature of those approaches is that both tend to exclude from consideration the functions that are *too wiggly*.

* Corresponding address: Instituto de Matemática Aplicada del Litoral, Güemes No 3450, CP 3000, Santa Fe, Argentina.
E-mail addresses: lloppamela@gmail.com (P. Llop), liliana.forzani@gmail.com (L. Forzani), rfraiman@udesa.edu.ar (R. Fraiman).

1 This is a part of the author Doctoral Dissertation.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.08.002

We, however, do not want to exclude *wiggly* curves since they allow us to obtain parametric rates of convergence. In practice, in our setting, when the data come in a discretized way, we just deal with the data vectors without any further pre-processing.

The irregular curve setting was exploited by Castellana and Leadbetter [8] who pointed, for Gaussian processes, that irregularity of trajectories corresponds to less correlation among $X(t)$ and $X(t+s)$ and therefore the trajectories contain more information that allows obtaining better rates of convergence of the variance to zero.

The problem we address is the nonparametric density estimation using k-nearest neighbors estimators, k-NN from now on, in the context of functional data. Given n-independent trajectories of a continuous time stochastic process verifying

$$X(t) = \mu(t) + e(t), \quad t \in T, \tag{1}$$

we will estimate the (marginal) density of $X(t)$. Here $\mu(t)$ stands for the mean function, and $e(t)$ is a zero mean, first-order stationary stochastic process with unknown marginal density function f_e.

The problem of estimating the density function using k-NN in $\mathbb{R}^d$ has been considered by several authors in different setups. In the independent case, Loftsgaarden and Quesenberry [29] were the first to prove its consistency. Since then, several results have been obtained. For instance, Moore and Henrichon [31] showed the uniform convergence, Wagner [41] proved the almost sure pointwise convergence and Boente and Fraiman [6] showed uniform consistency in the dependent case.

In the context of functional data (more precisely, when in the model (1) $\mu(t) = \mu$ is a constant independent of t), the problem of estimating the marginal density function when a single sample path is observed continuously over $[0, T]$ has been studied first by Castellana and Leadbetter [8] who showed that for irregular continuous time processes a parametric speed of convergence $\sqrt{T}$ is attained by δ-sequence-type density estimators. Some years later, Rosenblatt [37] showed the consistency of a kernel estimator for stationary Markov sequences and then, Nguyen [32] studied the almost sure convergence of a general class of recursive estimates of the density function in a continuous-time stationary Markov process. More recently, Blanke and Bosk [5] specified necessary conditions for getting parametric rates of convergence for δ-sequence estimators (although they make use of the kernel estimator in order to give explicit convergence rates), Blanke [4] obtained rates of pointwise and uniform almost sure convergence for adaptive estimators and Kutoyants [25] presented a review of several results concerning invariant density estimation for the ergodic diffusion process.

Labrador [26,27] defined and studied the k-NN estimator defined via local time in the same setting described above, i.e. when a single sample path is observed continuously over $[0, T]$ and proved convergence rates of order $\sqrt{\frac{T}{\log T}}$.

In this paper we consider that we have n-independent trajectories of continuous time stochastic processes verifying (1) with $\mu(t)$ not necessary a constant. Using the k-NN estimator given by Labrador for n-independent curves, we prove $\sqrt{n}$ consistency for μ constant since we get $\sqrt{n}$ asymptotic normality when the curves are irregular (see Section 2). Furthermore, we get strong rates of convergence when $\mu(t)$ is no longer a constant function (see Section 3). In Section 4 we apply our results to obtain a new classification rule for FDA. Moreover, some small simulations studies are presented. All proofs are given in the Appendix.

2. Estimating the density function of stationary processes

In this section, we estimate the marginal density function of the zero mean stationary stochastic process $e(t)$.

2.1. Preliminaries

Let $T \subset \mathbb{R}$ a finite interval, and $\{e(t) : t \in T\}$ a zero mean real-valued measurable continuous time stationary process with continuous trajectories defined on a rich enough probability space $(\Omega, \mathcal{A}, P)$. Let us suppose that $e(t)$ is a first-order stationary stochastic process with unknown marginal density function f_e which admits a local time. More precisely, we define the occupation measure associated with the process $e(t)$ as

$$\lambda(A, e) \doteq \lambda(A, e(t, \omega)) = \int_T \mathbb{I}_A(e(t, \omega))\mathrm{d}t, \quad A \in \mathcal{B}(\mathbb{R}),\ \omega \in \Omega.$$

Here, $\mathcal{B}(\mathbb{R})$ stands for the Borel sigma-algebra on $\mathbb{R}$. If λ is absolutely continuous with respect to the Lebesgue measure, then the local time is defined as a regular version of the Radon–Nikodym derivative $l_T(\cdot, e) \doteq l_T(\cdot, e(t, \omega))$ for almost all ω and therefore

$$\lambda(A, e) = \int_A l_T(u, e)\mathrm{d}u.$$

Let $\{e_1(t), \dots, e_n(t)\}$ a random sample of $e(t)$ and $I_{(x,r)} = [x-r, x+r]$. For $\{k_n\}$, $k_n/n < |T|$ a positive real numbers sequence converging to infinity, we define the random variable $H_n^e \doteq H_n^e(x)$ such that $\{e_1(t), \dots, e_n(t)\}$ spend time k_n at $I_{(x,H_n^e(x))}$. That is,

$$k_n = \sum_{i=1}^{n} \lambda(I_{(x,H_n^e(x))}, e_i) = \sum_{i=1}^{n} m\left(\{t \in T : |e_i(t) - x| \le H_n^e(x)\}\right), \tag{2}$$

where m stands for the Lebesgue measure in $\mathbb{R}$. We define the estimator for the density f_e as

$$\hat{f}_e(x) \doteq \frac{k_n}{2n|T|H_n^e(x)}, \tag{3}$$

where $|T|$ is the length of the interval T.

Remark 1. Observe that if the process $e(t)$ admits a local time, then $\hat{f}_e$ is well defined since H_n^e exists and it is unique. Indeed, if we define the function

$$G(r) \doteq \frac{1}{n}\sum_{i=1}^{n}\int_{I_{(x,r)}} l_T(u, e_i)\mathrm{d}u,$$

for n and x fixed, we have that $G(r)$ is a strictly increasing function of r with $G(0) = 0$. On the other hand, due to the existence of local time we can write

$$G(r) = \frac{1}{n}\sum_{i=1}^{n}\lambda(I_{(x,r)}, e_i),$$

then, $G(r) \to |T|$ when $r \to \infty$ and therefore, the existence and uniqueness of $H_n^e(x)$ is ensured. For a further reading on local times see [18].

2.2. General assumptions

We will consider the following assumptions:

H1 The sequence $\{e_i(t), 1 \le i \le n, t \in T\}$ are i.i.d. random elements with the same distribution as the stochastic process $\{e(t) : t \in T\}$ which admits a local time.
H2 $e(t)$ is a first-order stationary stochastic process with unknown density function f_e.
H3 $\{k_n\}$ is a positive real numbers sequence converging to infinity such that $k_n/n = o(1)$ and $\sum_{n=1}^{\infty}\exp(-ck_n) < \infty$ for each $c > 0$.
H4 The density f_e is a Lipschitz function.
H5 For each $c > 0$,

$$c^{-2}c_n^{-2}\int_T\int_T\int_{\{x:|u-x|\le cc_n\}}\int_{\{x:|v-x|\le cc_n\}}(f_{st}(u, v) - f_e(u)f_e(v))\mathrm{d}u\mathrm{d}v\mathrm{d}s\mathrm{d}t$$
$$\to \int_{T\times T}(f_{st}(x, x) - f_e^2(x))\mathrm{d}s\mathrm{d}t \doteq c_0^2(x) > 0,$$

where $c_n = \frac{k_n}{n}$, and f_{st} is the joint density of $(e(s), e(t))$.

Remark 2. A sufficient condition for H5 to hold is that there exist an integrable function $\psi(\cdot)$ such that

$$|f_{st}(u, v) - f_e(u)f_e(v)| \le \psi(t - s). \tag{4}$$

This follows from the Lebesgue-dominated convergence theorem together with the Lebesgue differentiation theorem.

Remark 3 (*Some comments on the assumptions*)**.** In assumption H1, the existence of the local time is required in order that the estimator $\hat{f}_e$ be well defined (see Remark 1). For conditions for the existence of local time we refer to Geman and Horowitz [18] and Karatzas and Shreve [24]. The conditions in H3 are exactly the same ones required to the smoothing parameter k_n in the classical finite dimensional setting for k-NN density estimates. H4 is also an standard condition to deal with the bias term in the finite dimensional case, and it is fulfilled for a large class of stochastic processes. Finally, H5 together with H1 are related to the irregularity of the trajectories. Indeed, as pointed in [8], condition (4) "is a strong dependence limitation between $e(s)$ and $e(t)$, when $t - s \to 0$, a feature that does not have a discrete time analog". In particular, they showed that for zero mean stationary Gaussian process with covariance function

$$\mathrm{cov}(e(s), e(t)) = 1 - C|t - s|^{\alpha} + o(|t - s|^{\alpha}), \quad C > 0,$$

when $0 < \alpha < 2$ condition (4) is fulfilled with $\psi(t - s) = 1 + K|t - s|^{-\alpha/2}$ in a neighborhood of $t - s = 0$ and $\psi(t - s) = K'|t - s|$ outside a neighborhood of $t - s = 0$ if the covariance is bounded away from 1 and integrable. Here, K and K' are constants. It is well known that Gaussian processes with $\alpha < 2$ have irregular trajectories in contrast to the more regular case $\alpha = 2$. Indeed, if a centered stationary Gaussian process have a covariance function given by

$$\mathrm{cov}(e(s), e(t)) = 1 - \frac{1}{2}C|t - s|^2 + O\left(\frac{(t - s)^2}{|\log|t - s||^a}\right), \quad C > 0,\ a > 3,$$

then, there exist a version of the process with a path of class C^1. See for instance, Example 1.2 and Proposition 1.11 in [1].

2.3. Consistency, strong convergence rates and asymptotic distribution

In Theorem 1 we prove, under mild conditions, the complete convergence of the estimator of the density f_e. Under some additional assumptions we obtain in Theorem 2 sharp bounds for strong rates of convergence and asymptotic normality with parametric rates of convergence in Theorem 3. The assumptions in these last theorems are closely related to those given in [8] where it is shown that for irregular continuous time processes a parametric speed of convergence is attained by kernel-type density estimators.

Theorem 1 (*Strong consistency*)**.** *Under* H1–H3, *for almost all* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \hat{f}_e(x) = f_e(x) \quad \text{a.co.}$$

Here "a.co." stands for the (almost) complete convergence.

Remark 4. If f_e is Lipschitz, the convergence is for all $x \in \mathbb{R}$.

Theorem 2 (*Strong rates of convergence*)**.** *Let us suppose that* H1, H2 *and* H4 *holds. Choose two sequences* $\{k_n\}$ *and* $\{v_n\}$ *of positive real numbers converging to infinity such that* $(k_n/n)\, v_n = o(1)$ *and* $\sum_{n=1}^{\infty} \exp\left(-c\,(k_n/v_n)\right) < \infty$ *for each* $c > 0$. *For that* k_n *let us also suppose that* H5 *holds. Then, for all* $x \in \mathbb{R}$,

$$\lim_{n\to\infty} v_n(\hat{f}_e(x) - f_e(x)) = 0 \quad \text{a.co.}$$

Remark 5. Our assumptions imply that we can choose k_n such that $v_n = n^{\gamma}$ for any $\gamma < \frac{1}{2}$, that is the type of result one gets for a parametric problem. Indeed, if $k_n = n^{\beta}$ and $v_n = n^{\gamma}$, in order that conditions $k_n/n = o(1)$ and $(k_n/n)\, v_n = o(1)$ hold, it suffices that $\beta < 1$ and $\beta - 1 + \gamma < 0$. Then, for any $\gamma < 1/2$, that is, $\gamma = 1/2 - \epsilon$ for some $0 < \epsilon < 1/2$, we can choose $\beta = 1/2 + \epsilon/2 < 1$ so that $\beta - 1 + \gamma < 0$.

Theorem 3 (*Asymptotic normality*)**.** *Let us assume that* H1 *and* H2 *holds and that the density* f_e *has two bounded derivatives. Choose a sequence* k_n *of positive real numbers such that* $\sqrt{n}/k_n = o(1)$ *and* $k_n/n^{3/4} = o(1)$. *For that* k_n, *let us also suppose that* H5 *holds. Then, for all* $x \in \mathbb{R}$,

$$\sqrt{n}(\hat{f}_e(x) - f_e(x)) \to \mathcal{N}\left(0, \frac{2|T|}{c_0(x)}\right).$$

Remark 6. Let us observe that if in model (1) $\mu(t)$ is a constant independent of t, $X(t)$ inherits all the properties of $e(t)$. This means that $X(t)$ will be a order one stationary process which admits a local time, so its density estimator, $\hat{f}_X$ from now on, will be computed in the same way as $\hat{f}_e$ and the results given in this section will be still true. However, it is clear that this is not the case if $\mu(t)$ is not constant. We consider this problem in the next section.

3. Estimating the density function of non-stationary processes

In this section, we extend the estimator given in Section 2 to a particular family of non-stationary stochastic processes. Suppose that in model (1) the deterministic mean function $\mu(t)$ is not constant with respect to the time, that is, suppose $X(t)$ is given by

$$X(t) = \mu(t) + e(t),$$

where $\mu(t)$ is a continuous function and $e(t)$ is a zero mean, first-order stationary stochastic process with unknown density function f_e. The density function of $X(t)$ will be denoted as $f_{X(t)} \doteq f_{X_t}$.

Let $\{X_1(t), \ldots, X_n(t)\}$ be independent trajectories with the same distribution as $X(t)$. We define the estimator of the density function f_{X_t} as

$$\hat{f}_{X_t}(x) = \hat{f}_u(x - \bar{X}_n(t)), \tag{5}$$

where

$$\hat{f}_u(x) \doteq \frac{k_n}{2n|T|H_n^u(x)}$$

with $u = \{U_{n1}, \ldots, U_{nn}\}$ given by

$$U_{ni}(t) = X_i(t) - \bar{X}_n(t) = e_i(t) - \bar{e}_n(t). \tag{6}$$

Here $\{e_1(t), \ldots, e_n(t)\}$ is a random sample of $e(t)$, $\bar{e}_n(t) = \frac{1}{n}\sum_{i=1}^{n} e_i(t)$ and H_n^u is defined as in (2) by replacing $\{e_1(t), \ldots, e_n(t)\}$ by u.

For any fixed t, the random variables $\{U_{n1}(t), \ldots, U_{nn}(t)\}$ are identically distributed but not necessarily independent. Therefore, we cannot use directly the results proved in Section 2.3. However we can still prove the complete convergence of the estimator of f_{X_t} and obtain strong rates of convergence.

Theorem 4. *For fixed t, let us suppose that* H1, H2 *and* H4 *holds. Choose two sequences $\{k_n\}$ and $\{v_n\}$ of positive real numbers converging to infinity such that $(k_n/n)\, v_n = o(1)$, $\sum_{n=1}^{\infty} \exp(-c\,(k_n/v_n)) < \infty$ for each $c > 0$ and $v_n(n/k_n)|\bar{e}_n(t)| \to 0$ a.co. Let us also suppose that the sequence k_n fulfils* H5. *Then, for all $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} v_n \left(\hat{f}_{X_t}(x) - f_{X_t}(x)\right) = 0 \quad \text{a.co.}$$

Remark 7. In this case our assumptions imply that we can choose k_n such that $v_n = n^\gamma$ for any $\gamma < \frac{1}{4}$. Indeed, let $k_n = n^\beta$ and $v_n = n^\gamma$. By Billingsley [3] we know that $\bar{e}_n(t) = o(n^{-\alpha})$ with $\alpha < 1/2$. In order that the conditions $v_n \frac{k_n}{n} \to 0$ and $v_n \frac{n}{k_n} |\bar{e}_n(t)| \to 0$ hold it suffices that $0 < \gamma < 1 - \beta$ and $0 < \gamma < \beta - \frac{1}{2}$, and γ is maximized under these two restrictions at $\gamma = \frac{1}{4}$.

Remark 8. In both cases, we compute the number of "nearest-neighbors" k_n by the standard leave-one-out cross-validation method.

4. A new classification rule for functional data

In this section, we apply our estimation results to obtain a new classification rule for functional data. The main aim in pattern recognition or supervised classification problems is to classify individuals into groups. Information about these groups is provided by a training sample $\{(X_i, Y_i) : 1 \le i \le n\}$, where each curve X_i has a label Y_i attached, indicating which group it belongs to. A new observation X is given without its label and we want to predict the unknown label.

The classical books by Devroye et al. [13], Duda and Stork [14] and Hastie et al. [22] provide a broad coverage of these topics, for the standard multivariate case where the variable $\mathbf{X}$ takes values in $\mathbb{R}^d$.

However, the definitions are mainly the same for an arbitrary metric space E. Given a finite set $\{1, \ldots, m\}$ and a metric space E, an observation is a pair $(x, y) \in E \times \{1, \ldots, m\}$, where x is known and y is a class or label that denotes the unknown nature of the observation. A mapping $g : E \to \{1, \ldots, m\}$ is called a classifier and represents our guess of the class y given its associated element $x \in E$. The classification is wrong if given an observation (x, y), $g(x) \neq y$.

Let $(X, Y) \in E \times \{1, \ldots, m\}$ be a random pair. Since an error occurs if $g(X) \neq Y$, the probability of misclassification for g is

$$L(g) = P(g(X) \neq Y). \tag{7}$$

The best possible classifier is the function g^* that minimizes (7). The minimum error probability (the Bayes error) is denoted by $L^* = L(g^*)$.

To obtain g^*, the distribution of (X, Y) should be known, but this is not typically the case. One must build up a classifier based on a training sample of independent pairs $\{(X_i, Y_i);\ 1 \le i \le n\}$, with the same distribution as the pair (X, Y) and known $Y_1, \ldots, Y_n$ values. Then a classifier is a function

$$g_n(\cdot; X_1, Y_1, \ldots, X_n, Y_n) : E \times (E \times \{1, \ldots, m\})^n \to \{1, \ldots, m\},$$

with probability of misclassification given by the conditional error probability

$$L_n(g_n) = P(g_n(X; X_1, Y_1, \ldots, X_n, Y_n) \neq Y | X_1, Y_1, \ldots, X_n, Y_n).$$

A sequence of classifiers $\{g_n;\ n \ge 1\}$ is called a rule.

In the finite dimensional case, there are several universally consistent classification rules. An important difference between the finite and the infinite-dimensional situations arises, however, with regard to consistency. Stone [40] provide general results for the universal consistency of a wide class of nonparametric classification rules, which in particular imply the consistency of most classical nonparametric rules. In particular, the most popular k-NN classifiers are *universally* (*weakly*) *consistent*, provided that $k = k_n \to \infty$ and $k/n \to 0$, as $n \to \infty$. This means that, for these k-NN classifiers, $L_n \to L^*$, in probability, as $n \to \infty$ (or equivalently $E(L_n) \to L^*$), *with no restriction at all on the underlying distribution* of $(\mathbf{X}, Y)$. In the infinite-dimensional case, this result is no longer true. This has been pointed out by Cèrou and Guyader [9] who have studied the consistency of the k-NN classifier when E is a metric space. Some recent results regarding supervised classification methods for functional data can be found in [11,12,9,13,14,19,23,28,30,33,38,7,10].

4.1. The rule

Let $E = C(T, \|\cdot\|)$ be the space of continuous functions on T, and $\|\cdot\|$ a norm on E. We want to classify a new data $X(t) \in E$ into one of the m classes $\mathcal{F}^j$, $j \in \mathcal{J} = \{1, \ldots, m\}$ using a training sample $\{(X_i(t), Y_i);\ 1 \le i \le n\}$ of i.i.d. random elements with the same distribution as the pair $(X(t), Y)$ and known $Y_1, \ldots, Y_n$ labels. We will assume that for each population $\mathcal{F}^j$ the model (1) holds; that is

$$X^j(t) = \mu^j(t) + e^j(t),$$

where $\mu^j(t)$ stands for the mean function of the population $\mathcal{F}^j, j \in \mathcal{J}$ and $e^j(t)$ is a zero mean, first-order stationary stochastic process with density unknown function f_e^j.

For each t fixed, the Bayes rule chooses the class $\mathcal{F}^j$ if and only if

$$f_e^j(X(t) - \mu^j(t)) > f_e^k(X(t) - \mu^k(t)), \quad \forall k \neq j.$$

This motivates defining our classification rule in the following way: we will classify $X(t)$ into the class $\mathcal{F}^j, j \in \mathcal{J}$ (and define $\widehat{Y} = j$) if and only if

$$m\left(\left\{t : \hat{f}_{X_t}^j(X(t)) > \hat{f}_{X_t}^k(X(t))\right\}\right) > m\left(\left\{t : \hat{f}_{X_t}^j(X(t)) \leq \hat{f}_{X_t}^k(X(t))\right\}\right), \quad \forall k \neq j, \tag{8}$$

where $\hat{f}_{X_t}^j$ is the estimator of $f_{X_t}^j$.

5. Some simulation studies

In order to illustrate the use of our estimation method, in this section, we perform some simulation studies of nonparametric functional density estimation and functional discrimination.

We built two samples from the original data set, the *learning sample* $(X_i(t), Y_i)_{i\in\mathcal{L}}$ and the *testing sample* $(X_j(t), Y_j)_{j\in\mathcal{T}}$. With the learning sample, we compute the density estimator for each group $(\hat{f}^1, \ldots, \hat{f}^m)$ using the cross-validated value $\hat{k}_n^j, j = 1, \ldots, m$. In order to measure the discriminant power of our method, we evaluate the estimators obtained with the learning samples at the testing sample and we classify it according to the rule given by (8). Finally, we compute the *misclassification rate* as

$$\text{Misclas} = \frac{1}{\sharp\mathcal{T}} \sum_{j\in\mathcal{T}} \mathbb{I}_{\{\widehat{Y}_j \neq Y_j\}}.$$

Example 1. We start testing our method with a very simple model that fulfills all the assumptions. We consider a stochastic process $X(t)$ defined by

$$X(t) = \mu + \sigma e(t), \quad t \in T = (0, 1), \tag{9}$$

where

$$e(t) = \frac{w(t)}{\sqrt{t}}, \quad \text{with } w(t) \text{ the standard Wiener process.}$$

Remark 9. Since μ is constant, $X(t)$ as the same properties as $e(t)$ and therefore $\hat{f}_X$ will be computed with the tools given in Section 2 for computing $\hat{f}_e$ (see Remark 6).

In the first stage, we consider $\mu = 0$ and $\sigma = 1$ so that $X(t)$ is stationary and, for each t, $X(t) \sim \mathcal{N}(0, 1)$. Fig. 1 shows the theoretical density function of $X(t)$ and its estimator computed from a sample of size 200 measured at 100 equally spaced points on [0, 1]. As we can see in Fig. 1, the estimator fits very well the true density except in the tails where, due to the nature of the processes, we have not enough data to perform a good estimation. To assess the performance of our classification method, in the second stage we considered two classes under the model (9), both with $\sigma = 1$ but one of them with μ constant and equal to 0, and the another one with mean $\mu \neq 0$. In particular, we will consider the cases where $\mu = 0.5, 1.5, 2.5, 3.5$. We generate a learning sample of size 200 (100 of each class) measured at 150 instants of time in the interval [0, 1] and a testing sample of the same size. With the learning sample we compute the estimators for each class and then, in order to obtain the misclassification error, we evaluate them in the testing sample.

We repeat this procedure 50 times in order to obtain 50 misclassification rates for each case which are shown in Fig. 2. Let us note how the Bayes errors get smaller as the means goes far apart; this is due to the fact that when we classify two populations which are very close in mean their densities present a considerable amount of overlap, making it difficult to distinguish between groups (see Fig. 3).

Example 2. Recently, Shin [39] proposed an extension of Fisher discriminant analysis for stochastic processes, refer to InfFLD, which uses a bijective mapping that connects a second-order stationary process with the reproducing kernel Hilbert space generated by its within class covariance kernel. In particular, he provides the results of a simulation study comparing the InfFLD method with the classical multivariate Fisher's (FLD), penalized discriminant analysis (PDA) using both the ridge penalty (PDA/Ridge) and a penalty matrix for cubic spline smoothing (PDA/Spline) principal components analysis (NPCD/PCA) and multivariate partial least-squares regression (NPCD/MPLSR) (nonparametric curve discrimination methods proposed by Ferraty and Vieu [16]).

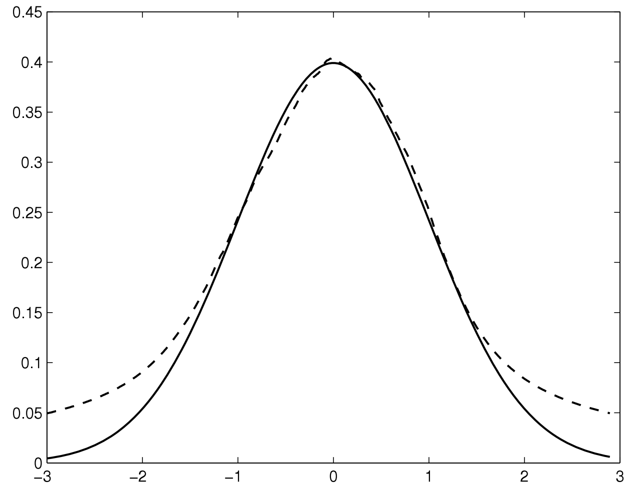

Fig. 1. Estimated (dashed curve) and theoretical (solid curve) density functions of $X(t)$ for $k_n = 43.196$.

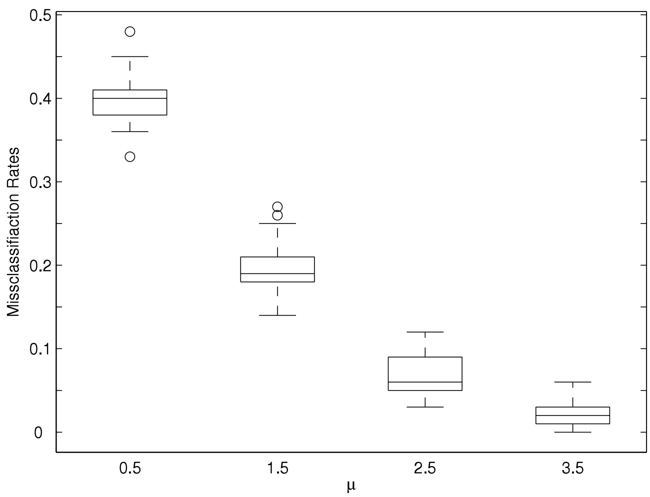

Fig. 2. Boxplot of the misclassification error from 50 runs.

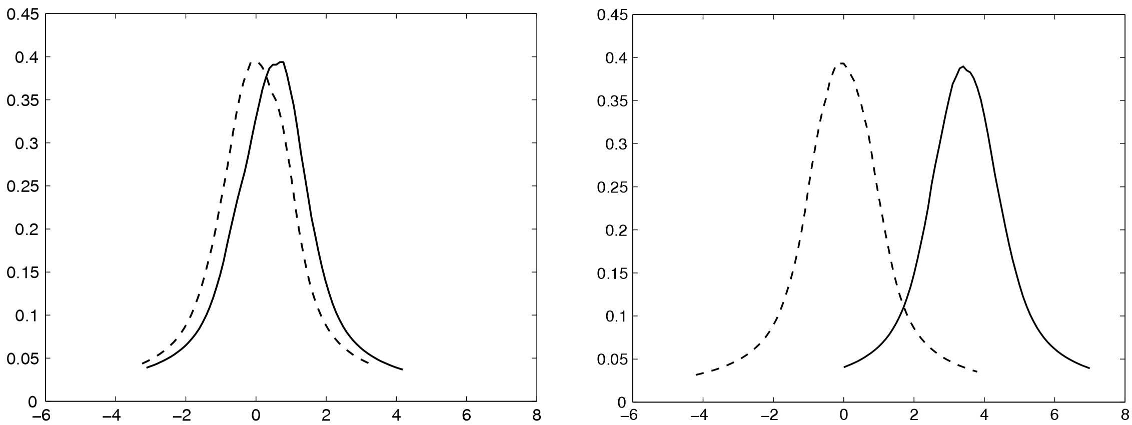

Fig. 3. (a) Density estimator for a $\mathcal{N}(0, 1)$ (dashed curve) and for a $\mathcal{N}(0.5, 1)$ (solid curve). (b) Density estimator for a $\mathcal{N}(0, 1)$ (dashed curve) and for a $\mathcal{N}(3.5, 1)$ (solid curve).

In those simulations, the populations were generated following the stochastic processes

$$X_1(t) = 3\sqrt{2}\cos(\pi t) + \sqrt{2}\cos(2\pi t) + e(t) \quad \text{and} \quad X_2(t) = \sqrt{2}\cos(2\pi t) + e(t),$$

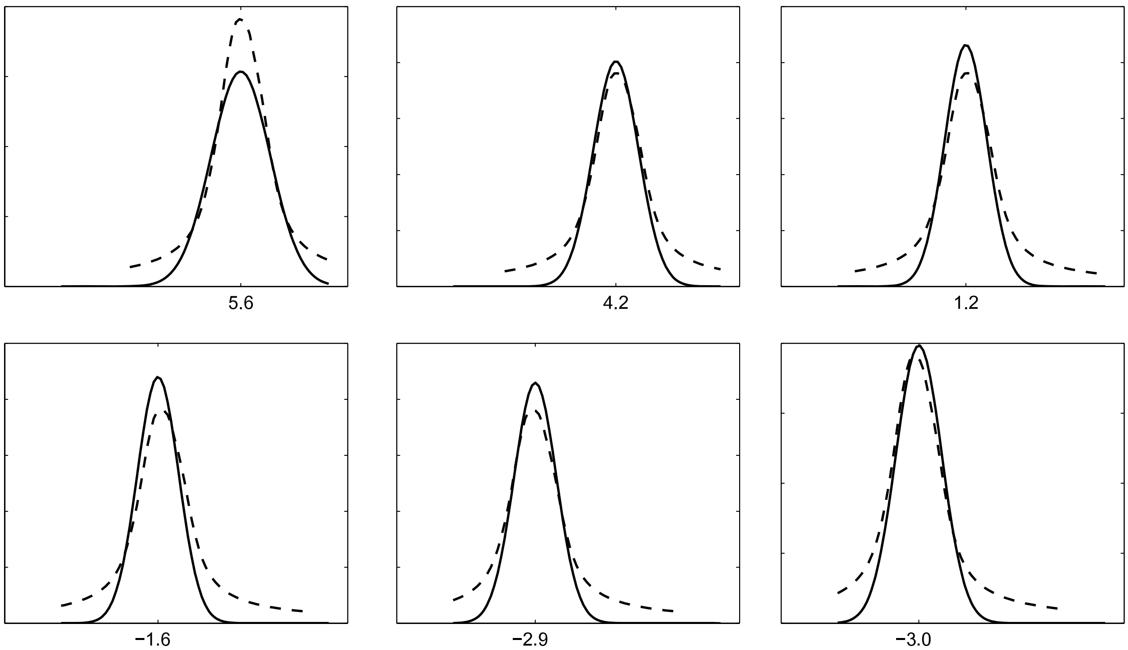

Fig. 4. Estimated (dashed curve) and theoretical (solid curve) density functions of $X_1(t)$ for $t = 0.013, 0.180, 0.347, 0.513, 0.680, 0.847$.

with

$$e(t) = \sum_{i=1}^{30} i^{-1/2} U_i \sqrt{2}\cos(i\pi t),$$

where U_i are i.i.d. standard normal random variables. Let us observe that for each t,

$$X_1(t) \sim \mathcal{N}\left(3\sqrt{2}\cos(\pi t) + \sqrt{2}\cos(2\pi t), 2\sum_{i=1}^{30}\cos^2(\pi ti)/i\right)$$

and

$$X_2(t) \sim \mathcal{N}\left(\sqrt{2}\cos(2\pi t), 2\sum_{i=1}^{30}\cos^2(\pi ti)/i\right).$$

For each class, he generates a learning sample of size 100 (50 of each class) measured at 100 instants of time in the interval [0, 1] and a testing sample of size 500. For our proposal, we run the same experiment with the same size of learning and testing samples. Figs. 4 and 5 show, respectively, the density estimator and the theoretical density function of $X_1(t)$ and $X_2(t)$ for some instants of time, the dashed curves correspond to the density estimators and the solid curves correspond to the theoretical density function. As in the stationary case (Example 1), the density estimator fits very well the true density except in the tails where we do not have enough information. Next, we evaluate this estimator in the testing sample in order to obtain the misclassification error. This procedure was replicated an additional 49 times by randomly building 49 learning samples and 49 testing samples. Finally, we get 50 misclassification rates.

In Table 1 we reproduce the results given in [39] and add the result obtained with our method, referred to as NPDE. As we can see, the NPDE method behaves slightly worse than InfFLD but better than the other methods.

Let us observe that $e(t)$ is a non-stationary process since its variance (and consequently their distribution) depends on the time. This shows that our classification method is robust with respect to non-stationarity.

6. Conclusions

In this paper we have proposed a nonparametric density estimation method for functional data following the model

$$X(t) = \mu(t) + e(t), \quad t \in T,$$

where $\mu(t)$ stands for the deterministic mean function, and $e(t)$ is a zero mean, first-order stationary stochastic process which admits a local time with unknown density function f_e.

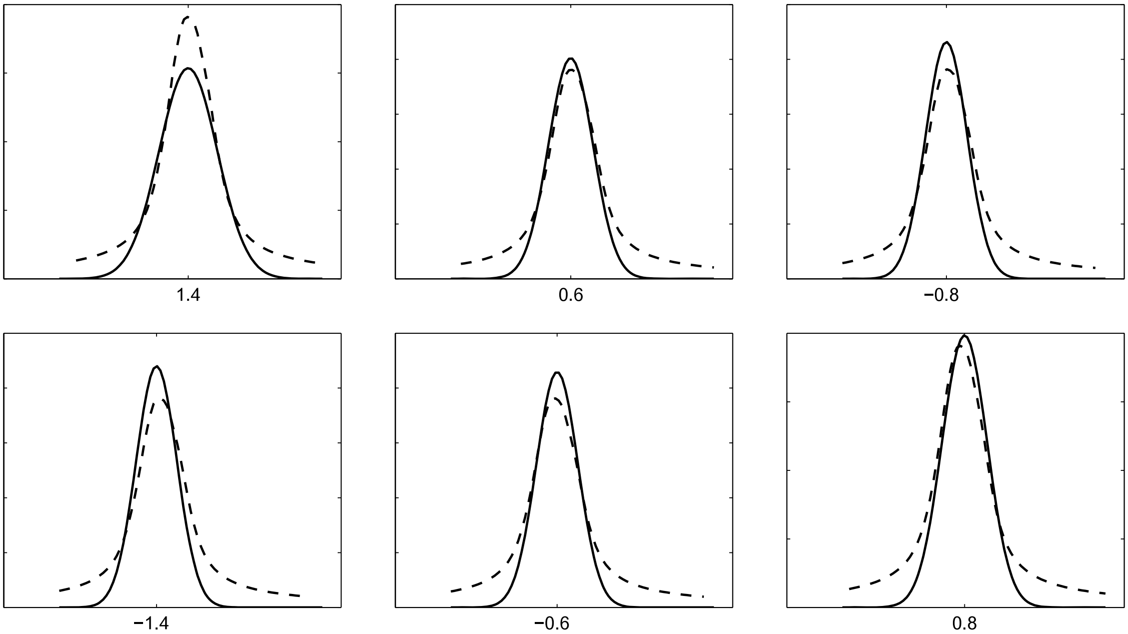

Fig. 5. Estimated (dashed curve) and theoretical (solid curve) density functions of $X_2(t)$ for $t = 0.013, 0.180, 0.347, 0.513, 0.680, 0.847$.

Table 1
Mean, median and standard deviation of the miss-classification errors for the RKHS-based algorithm (InfFLD), classical Fisher's method (FLD), penalized discriminant method (PDA), nonparametric curve discrimination method (MPLSR) and nonparametric density estimation (NPDE).

Method	Mean	Median	Standard error
InfFLD	0.0832	0.082	0.0109
FLD	0.2086	0.2	0.0418
PDA/Ridge	0.0889	0.086	0.0181
PDA/Spline	0.0891	0.087	0.0163
NPCD/PCA	0.0906	0.086	0.0152
NPCD/MPLSR	0.0992	0.089	0.0323
NPDE	0.0889	0.0890	0.0096

First, we obtained an estimator for the marginal density function of $e(t)$, which is the same for all t. We show that it is strongly consistent with rate of convergence n^{α}, for any $\alpha < 1/2$ and that it has asymptotic normal distribution with rate $\sqrt{n}$. If $\mu(t) = \mu$ is constant, $X(t)$ inherits the properties of $e(t)$ so, in the same way as for $e(t)$, we can compute the estimator of f_X. Though this is not new in nonparametric setting (see [8]), it is a surprising and desired property.

When $\mu(t)$ is nonconstant, $X(t)$ does not inherit the stationarity of $e(t)$ and therefore it has a different marginal density function for each t. In this context, the estimator has shown to be strongly consistent for each t with a smaller convergence rate than that in the stationary case.

In simulations studies, we computed the density estimator and we applied the estimation results to obtain a new classification rule for functional data.

Appendix

Proof of Theorem 1. Let

$$C_n = \left\{ \left| \hat{f}_e(x) - f_e(x) \right| > \epsilon \right\}.$$

By definition of complete convergence we need to show that $\sum_{n=1}^{\infty} P(C_n) < \infty$ for all $\epsilon > 0$. Let x be fixed so that the Lebesgue differentiation theorem holds for f_e in x. For this x, using the decomposition introduced by Wagner [41], we can write

$$C_n = A_n \cup B_n,$$

with

$$A_n = \left\{ H_n(x) < \frac{k_n}{2n|T|(f_e(x) + \epsilon)} \right\}$$

and

$$B_n = \begin{cases} \left\{ H_n(x) > \dfrac{k_n}{2n|T|(f_e(x)-\epsilon)} \right\} & \text{if } f_e(x) > \epsilon \\ \emptyset & \text{if } f_e(x) \le \epsilon. \end{cases}$$

It will be sufficient to show that

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} P(B_n) < \infty. \tag{A.1}$$

The proof of the right-side inequality of (A.1) is similar to the one of the left-side inequality and therefore it will be omitted. In order to prove the left-side inequality of (A.1), let us define $a_n = \frac{k_n}{2n|T|(f_e(x)+\epsilon)}$. Then, we have

$$H_n < a_n \Leftrightarrow \sum_{i=1}^{n} \underbrace{\int_T \mathbb{I}_{I(x,a_n)}(e_i(t))\mathrm{d}t}_{\doteq Y_{ni}} > k_n,$$

where Y_{ni} are independent random variables and

$$P(A_n) = P\left(\sum_{i=1}^{n} Y_{ni} > k_n\right).$$

Let $p_n = \int_{\{u:|u-x|\le a_n\}} f_e(u)\mathrm{d}u$. Since the marginal distributions of e are the same for each t, we obtain

$$\mathbb{E}(Y_{ni}) = \mathbb{E}\left(\int_T \mathbb{I}_{I(x,a_n)}(e_i(t))\mathrm{d}t\right) = \int_T P(e_i(t) \in I_{(x,a_n)})\mathrm{d}t = |T|p_n \tag{A.2}$$

and using Cauchy–Schwartz inequality

$$\mathbb{E}\left(Y_{ni}^2\right) = \mathbb{E}\left(\left(\int_T \mathbb{I}_{I(x,a_n)}(e_i(t))\mathrm{d}t\right)^2\right) \le |T|\mathbb{E}\left(\int_T \left(\mathbb{I}_{I(x,a_n)}(e_i(t))\right)^2 \mathrm{d}t\right) = |T|^2 p_n.$$

Let us define $\widetilde{Y}_{ni} \doteq Y_{ni} - \mathbb{E}(Y_{ni})$. Then, using (A.2), we obtain

$$\begin{aligned} \sum_{n=1}^{\infty} P(A_n) &= \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n - n|T|p_n\right) \\ &= \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n\left(1 - \frac{p_n}{2a_n(f_e(x)+\epsilon)}\right)\right); \end{aligned} \tag{A.3}$$

where in the last equality we have used the relation between a_n and k_n. Since $a_n \to 0$, by the Lebesgue differentiation theorem we have that $p_n/2a_n \to f_e(x)$. Therefore, there exists $N_1 = N_1(x)$, such that if $n \ge N_1(x)$

$$\left|\frac{p_n}{2a_n} - f_e(x)\right| < \epsilon/2 \Rightarrow 1 - \frac{p_n}{2a_n(f_e(x)+\epsilon)} > \frac{\epsilon}{2(f_e(x)+\epsilon)} = C(x,\epsilon) \doteq C. \tag{A.4}$$

With (A.4) in (A.3), using that $\left|\widetilde{Y}_{ni}\right| \le 2|T|$ and $\mathrm{var}\left(\widetilde{Y}_{ni}\right) \le |T|^2 p_n$, we apply the Bernstein inequality [2] and conclude that, for $n \ge N_1(x)$,

$$\begin{aligned} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n\left(1 - \frac{p_n}{2a_n(f_e(x)+\epsilon)}\right)\right) &\le P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n C\right) \\ &\le 2\exp\left(-\frac{k_n^2 C^2}{2n|T|^2 p_n + 4|T|k_n C}\right). \end{aligned} \tag{A.5}$$

In order to bound the exponent we use that (A.4) implies $p_n < (f_e(x)+\epsilon)2a_n = k_n/n|T|$ and as a consequence

$$\frac{k_n^2 C^2}{2n|T|^2 p_n + 4|T|k_n C} > \frac{k_n^2 C^2}{2|T|k_n + 4|T|k_n C} = k_n \frac{C^2}{2|T|(1+2C)}.$$

Replacing this bound into (A.5) we get, for $n \geq N_1(x)$, that

$$P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n\left(1 - \frac{p_n}{2a_n(f_e(x)+\epsilon)}\right)\right) \leq 2\exp\left(-ck_n\right),$$

with $c = \frac{C^2}{2|T|(1+2C)}$. Finally hypothesis H3 implies $\sum_{n=N_1(x)}^{\infty} \exp\left(-k_n c\right) < \infty$ and the theorem follows. □

Proof of Theorem 2. Let

$$C_n = \left\{ v_n \left|\hat{f}_e(x) - f_e(x)\right| > \epsilon \right\},$$

we need to prove that $\sum_{n=1}^{\infty} P(C_n) < \infty$ for all $\epsilon > 0$. We do the analysis analogous that in Theorem 1, replacing ϵ by $\epsilon_n = \frac{\epsilon}{v_n}$ and we get (A.3) for ϵ_n. That is,

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > k_n\left(1 - \frac{p_n}{2a_n(f_e(x)+\epsilon_n)}\right)\right). \tag{A.6}$$

The mean value theorem and the Lipschitz condition for f_e ensure the existence of $x_n \in I(x, a_n)$ for which $p_n/2a_n = f_e(x_n)$. Using this and the Lipschitz condition again we obtain

$$\left|\frac{p_n}{2a_n} - f_e(x)\right| = |f_e(x_n) - f_e(x)| \leq K|x_n - x| \leq Ka_n. \tag{A.7}$$

Now, by the definition of a_n, the fact that $\epsilon_n \to 0$ and the hypothesis $(k_n/n)v_n = o(1)$, for all $\epsilon > 0$ there exists N_1 such that, for $n \geq N_1(x)$,

$$Ka_n \leq C_1(x)\frac{k_n}{n} \leq \frac{1}{2}\frac{\epsilon}{v_n} = \frac{\epsilon_n}{2}. \tag{A.8}$$

Therefore from (A.7) and (A.8) we have

$$1 - \frac{1}{f_e(x)+\epsilon_n}\frac{p_n}{2a_n} \geq \frac{\epsilon_n - Ka_n}{f_e(x)+\epsilon_n} \geq \frac{C_2}{v_n}.$$

With this in (A.6),

$$\sum_{n=1}^{\infty} P(A_n) \leq \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > C_2\frac{k_n}{v_n}\right).$$

Now, since $a_n \sim k_n/n$, from H5 we obtain

$$\frac{1}{a_n^2}\operatorname{var}\left(\widetilde{Y}_{ni}\right) = \frac{1}{a_n^2}\int_T\int_T\int_{\{u:|u-x|\leq a_n\}}\int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_e(u)f_e(v))\,\mathrm{d}u\mathrm{d}v\mathrm{d}t\mathrm{d}s \to c_0^2 > 0, \tag{A.9}$$

then for $n \geq N_2(x)$, $\operatorname{var}\left(\widetilde{Y}_{ni}\right) \leq C_3 a_n^2$. Applying Bernstein inequality with $\left|\widetilde{Y}_{ni}\right| \leq 2|T|$, $\operatorname{var}\left(\widetilde{Y}_{ni}\right) \leq C_3 a_n^2$, we obtain for $n \geq N_3(x) = \max\{N_1(x), N_2(x)\}$

$$\begin{aligned}\sum_{n=1}^{\infty} P(A_n) &\leq \sum_{n=1}^{\infty} P\left(\sum_{i=1}^{n} \widetilde{Y}_{ni} > C_2\frac{k_n}{v_n}\right) \\ &\leq 2\exp\left\{-\left(\frac{C_3}{C_4\frac{k_n}{n} + C_5}\right)\frac{k_n}{v_n}\right\}.\end{aligned}$$

In order to bound the exponent we use the fact that $k_n/n \to 0$ and then we obtain

$$\sum_{n=1}^{\infty} P(A_n) \leq 2\sum_{n=1}^{\infty} \exp\left\{-C_6\frac{k_n}{v_n}\right\} < \infty.$$

Finally using that $\sum_{n=1}^{\infty} \exp\left(-c\left(k_n/v_n\right)\right) < \infty$, for each $c > 0$ we get the theorem. □

Proof of Theorem 3. We do the analysis analogous to Theorem 1 where we replace ϵ by $\epsilon_n = \frac{t}{\sqrt{n}}$ and we get calling $S_n = \sum_{i=1}^{n} Y_{ni}$ and $s_n^2 = \operatorname{var}(S_n)$

$$P\left(\sqrt{n}\left(\hat{f}_e(x) - f_e(x)\right) \leq t\right) \leq P\left(\frac{S_n - E(S_n)}{s_n} \leq \frac{k_n - n|T|p_n}{s_n}\right). \tag{A.10}$$

Now, since by (A.9) $s_n^2 = O(na_n^2)$, by Lindenberg theorem

$$\frac{S_n - E(S_n)}{s_n} \to N(0, 1), \tag{A.11}$$

where the convergence is in distribution. On the other hand,

$$\begin{aligned}\frac{k_n - n|T|p_n}{s_n} &= \frac{2na_n|T|(f_e(x) + t/\sqrt{n}) - n|T|\int_{x-a_n}^{x+a_n} f_e(u)\mathrm{d}u}{s_n} \\ &= s_n^{-1}n|T|\int_{x-a_n}^{x+a_n}(f_e(x) - f_e(u))\mathrm{d}u + \frac{2na_n|T|t}{s_n\sqrt{n}}.\end{aligned} \tag{A.12}$$

By Taylor theorem, there exists a number x^* between x and u such that

$$\int_{x-a_n}^{x+a_n}(f_e(x) - f_e(u))\mathrm{d}u = -\frac{1}{2}\int_{x-a_n}^{x+a_n} f_e''(x^*)(u-x)^2\mathrm{d}u.$$

Since f has two derivatives bounded

$$\left|s_n^{-1}n|T|\int_{x-a_n}^{x+a_n}(f_e(x) - f_e(u))\mathrm{d}u\right| \le Cs_n^{-1}na_n^3.$$

Therefore, in (A.12) we have

$$\frac{k_n - n|T|p_n}{s_n} = O(s_n^{-1}na_n^3) + \frac{2na_n|T|t}{s_n\sqrt{n}}.$$

Since $s_n^{-1}na_n^3 \to 0$ and by (A.9) $s_n^2/(na_n^2) \to c_o^2$,

$$\lim_{n\to\infty}\frac{k_n - E(S_n)}{s_n} = \frac{2|T|}{c_0}t.$$

Finally from this, (A.11) and (A.10) we get the result. □

Lemma 5. *Let suppose that* H1–H3 *holds and that for fixed t, $v_n(n/k_n)|\bar{e}_n(t)| = o(1)$ a.co. Then, for all $x \in \mathbb{R}$, we have that*

$$\lim_{n\to\infty} v_n\left(\hat{f}_u(x - \bar{X}_n(t)) - \hat{f}_e(x - \mu(t))\right) = 0, \quad \text{a.co.}$$

In order to prove this lemma we need an auxiliary result.

Lemma 6. *For fixed t, let H_n^u and $\bar{e}_n(t)$ as defined in (6) where $u = \{U_{n1}, \dots, U_{nn}\}$ with $U_{ni}(t) = e_i(t) - \bar{X}_n(t) = e_i(t) - \bar{e}_n(t)$ and H_n^e as defined in (2). Then, for each n, x,*

$$\left|H_n^u(x - \bar{X}_n(t)) - H_n^e(x - \mu(t))\right| \le 2|\bar{e}_n(t)|.$$

Proof of Lemma 6. It is an immediate consequence of

(i) $\left|H_n^u(x - \bar{X}_n(t)) - H_n^u(x - \mu(t))\right| \le |\bar{e}_n(t)|$;
(ii) $\left|H_n^u(x - \mu(t)) - H_n^e(x - \mu(t))\right| \le |\bar{e}_n(t)|$.

We will prove only (i) since the proof of (ii) is analogous. Let x fixed, using that $\bar{X}_n(t) = \mu(t) + \bar{e}_n(t)$ we obtain

$$\left\{t : \left|U_{ni}(t) - (x - \bar{X}_n(t))\right| \le H_n^u(x - \bar{X}_n(t))\right\} \subset \left\{t : |U_{ni}(t) - (x - \mu(t))| \le H_n^u(x - \bar{X}_n(t)) + |\bar{e}_n(t)|\right\};$$

therefore,

$$k_n = \sum_{i=1}^n\int_T \mathbb{I}_{\{|U_{ni}(t)-(x-\bar{X}_n(t))|\le H_n^u(x-\bar{X}_n(t))\}}(t)\mathrm{d}t \le \sum_{i=1}^n\int_T \mathbb{I}_{\{|U_{ni}(t)-(x-\mu(t))|\le H_n^u(x-\bar{X}_n(t))+|\bar{e}_n(t)|\}}(t)\mathrm{d}t,$$

and for the definition of k_n we obtain

$$H_n^u(x - \mu(t)) \le H_n^u(x - \bar{X}_n(t)) + |\bar{e}_n(t)|. \tag{A.13}$$

Similarly, we can prove that

$$H_n^u(x - \bar{X}_n(t)) \le H_n^u(x - \mu(t)) + |\bar{e}_n(t)|. \tag{A.14}$$

Finally, from (A.13) and (A.14) we have

$$\left|H_n^u(x-\bar{X}_n(t)) - H_n^u(x-\mu(t))\right| \le |\bar{e}_n(t)|. \quad \square$$

Proof of Lemma 5. Let $\epsilon > 0$, x, t fixed, $\hat{f}_e$ as defined in (3), then

$$\begin{aligned} \upsilon_n \left|\hat{f}_u(x-\bar{X}_n(t)) - \hat{f}_e(x-\mu(t))\right| &= \frac{k_n \upsilon_n}{2n|T|} \frac{\left|H_n^e(x-\mu(t)) - H_n^u(x-\bar{X}_n(t))\right|}{H_n^u(x-\bar{X}_n(t))H_n^e(x-\mu(t))} \\ &\le \frac{k_n \upsilon_n}{2n|T|} \frac{2|\bar{e}_n(t)|}{H_n^u(x-\bar{X}_n(t))H_n^e(x-\mu(t))} \end{aligned}$$

where in the last inequality we have used Lemma 6. By Theorem 1, $\hat{f}_e(x-\mu(t)) < f_e(x-\mu(t)) + \epsilon$ for all $n \ge N_1(x, \epsilon, t)$ which implies that

$$H_n^e(x-\mu(t)) > C_1(x, t, \epsilon)\frac{k_n}{n} = C_1\frac{k_n}{n}. \tag{A.15}$$

By Lemma 6 and (A.15)

$$H_n^u(x-\bar{X}_n(t)) + 2|\bar{e}_n(t)| \ge C_1\frac{k_n}{n}. \tag{A.16}$$

Since by hypothesis $\upsilon_n \frac{n}{k_n}|\bar{e}_n(t)| \to 0$, for all $n \ge N_2(x, \epsilon, t)$ we have that

$$\frac{n}{k_n}|\bar{e}_n(t)| \le \frac{1}{4}C_1 \quad \text{and therefore } C_1\frac{k_n}{n} - 2|\bar{e}_n(t)| \ge \frac{1}{2}C_1\frac{k_n}{n}.$$

Replacing in (A.16) we obtain

$$H_n^u(x-\bar{X}_n(t)) \ge C_1\frac{k_n}{n} - 2|\bar{e}_n(t)| \ge \frac{C_1}{2}\frac{k_n}{n}. \tag{A.17}$$

So from (A.15) and (A.17), for all $n \ge \max\{N_1, N_2\}$ we obtain

$$\begin{aligned} \upsilon_n \left|\hat{f}_u(x-\bar{X}_n(t)) - \hat{f}_e(x-\mu(t))\right| &< \frac{1}{|T|}\frac{k_n \upsilon_n}{n}\frac{|\bar{e}_n(t)|}{\frac{C_1}{2}\frac{k_n}{n}C_1\frac{k_n}{n}} \\ &= C_3 \upsilon_n \frac{n}{k_n}|\bar{e}_n(t)|. \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(\upsilon_n \left|\hat{f}_u(x-\bar{X}_n(t)) - \hat{f}_e(x-\mu(t))\right| \ge \epsilon\right) &\le \sum_{n=1}^{\infty} P\left(C_3 \upsilon_n \frac{n}{k_n}|\bar{e}_n(t)| \ge \epsilon\right) \\ &= \sum_{n=1}^{\infty} P\left(\upsilon_n \frac{n}{k_n}|\bar{e}_n(t)| \ge \frac{\epsilon}{C_3}\right) \\ &= \sum_{n=1}^{\infty} P\left(\upsilon_n \frac{n}{k_n}|\bar{e}_n(t)| \ge \epsilon_0\right) < \infty, \end{aligned}$$

which concludes the proof. $\square$

Proof of Theorem 4. By definition (5) and since $X(t)$ satisfies the model (1) we need to prove

$$\lim_{n\to\infty} \upsilon_n \left(\hat{f}_u(x-\bar{X}_n(t)) - f_e(x-\mu(t))\right) = 0, \quad \text{a.co.}$$

Let $\epsilon > 0$ and x, t fixed. Let observe that

$$\begin{aligned} P\left\{\upsilon_n \left|\hat{f}_u(x-\bar{X}_n(t)) - f_e(x-\mu(t))\right| \ge \epsilon\right\} &\le P\left\{\upsilon_n \left|\hat{f}_u(x-\bar{X}_n(t)) - \hat{f}_e(x-\mu(t))\right| \ge \epsilon/2\right\} \\ &\quad + P\left\{\upsilon_n \left|\hat{f}_e(x-\mu(t)) - f_e(x-\mu(t))\right| \ge \epsilon/2\right\}. \end{aligned}$$

The result follows from Lemma 5 and Theorem 2. $\square$

References

[1] J. Azaïs, M. Wchebor, Level Sets and Extrema or Random Processes and Fields, Wiley, 2009.
[2] S. Bernstein, Probability Theory, M.L. Gostechizdat, 1946.
[3] P. Billingsley, Convergence of Probability Masures, John Wiley, 1968.
[4] D. Blanke, Adaptive sampling schemes for density estimation, J. Statist. Plann. Inference 136 (9) (2004) 2898–2917.
[5] D. Blanke, D. Bosq, Accurate rates of density estimators for continuous-time processes, Statist. Probab. Lett. 33 (2) (1997) 185–191.
[6] G. Boente, R. Fraiman, Consistency of a nonparametric estimate of a density function for dependent variables, J. Multivariate Anal. 25 (1) (1988) 90–99.
[7] F. Burba, F. Ferraty, P. Vieu, k-nearest neighbor method in functional nonparametric regression, J. Nonparametr. Stat. 21 (4) (2009) 453–469.
[8] J.V. Castellana, M.R. Leadbetter, On smoothed probability density estimation for stationary processes, Stochastic Process. Appl. 21 (2) (1986) 179–193.
[9] F. Cèrou, A. Guyader, Nearest neighbor classification in infinite dimension, ESAIM Probab. Stat. 10 (2006) 340–355.
[10] D. Costanzo, C. Preda, G. Saporta, Anticiped prediction in discriminant analysis on functional data for binary response, Comput. Statist. 22 (2) (2006) 223–235.
[11] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Statist. 22 (3) (2007) 481–496.
[12] A. Cuevas, R. Fraiman, On depth measures and dual statistics. A methodology for dealing with general data, J. Multivariate Anal. 100 (4) (2009) 753–766.
[13] L. Devroye, L. Gyorfi, G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer-Verlag, 1996.
[14] P. Duda, R.O. Hart, D. Stork, Pattern Classification, 2nd ed., Wiley, 2000.
[15] F. Ferraty, High-dimensional data: a fascinating statistical challenge, J. Multivariate Anal. 101 (2010) 305–306.
[16] F. Ferraty, P. Vieu, Curves discrimination: a nonparametric functional approach, Comput. Statist. Data Anal. 44 (1–2) (2003) 161–173.
[17] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis. Theory and Practice, Springer, New York, 2006.
[18] D. Geman, J. Horowitz, Occupation densities, Ann. Probab. 8 (1) (1980) 1–67.
[19] A.K. Ghosh, P. Chaudhuri, On maximal depth and related classifiers, Scand. J. Statist. 32 (2005) 327–350.
[20] W. González Manteiga, P. Vieu, Statistics for functional data, Comput. Statist. Data Anal. 51 (2007) 4788–4792.
[21] T. Hastie, A. Buja, R. Tibshirani, Penalized discriminant analysis, Ann. Statist. 23 (1) (1995) 73–102.
[22] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer-Verlag, New York, 2001.
[23] G.M. James, T.J. Hastie, Functional linear discriminant analysis for irregularly sampled curves, J. R. Stat. Soc. Ser. B 63 (2001) 533–550.
[24] I. Karatzas, S. Shreve, Brownian Motion and Stochastic Calculus, Springer, 1998.
[25] Y. Kutoyants, On invariant density estimation for ergodic diffusion processes, SORT 28 (2) (2004) 111–124.
[26] B. Labrador, Almost sure convergence of the k_T-occupation time density estimator, C. R. Math. Acad. Sci. Paris 343 (10) (2006) 665–669.
[27] B. Labrador, Strong pointwise consistency of the k_T-occupation time density estimator, Statist. Probab. Lett. 78 (9) (2008) 1128–1137.
[28] B. Li, Q. Yu, Classification of functional data: a segmentation approach, Comput. Statist. Data Anal. 52 (10) (2008) 4790–4800.
[29] D. Loftsgaarden, C. Quesenberry, A nonparametric estimate of a multivariate density function, Ann. Math. Statist. 36 (3) (1965) 1049–1051.
[30] S. Lopez-Pintado, J. Romo, Depth based classification for functional data, DIMACS Ser. Discrete Math. 72 (2006) 103–120.
[31] D. Moore, E. Henrichon, Uniform consistency of some estimates of a density function, Ann. Math. Statist. 40 (4) (1969) 1499–1502.
[32] H. Nguyen, Density estimation in a continuous-time stationary Markov process, Ann. Statist. 7 (2) (1979) 341–348.
[33] C. Preda, G. Saporta, C. Lévéder, PLS classification of functional data, Comput. Statist. 22 (2) (2007) 223–235.
[34] J. Ramsay, C. Dalzell, Some tools for functional data analysis, J. R. Stat. Soc. Ser. B 53 (1991) 539–572.
[35] J. Ramsay, B. Silverman, Applied Functional Data Analysis, in: Springer Series in Statistics, Springer, New York, 2002.
[36] J. Ramsay, B. Silverman, Functional Data Analysis, in: Springer Series in Statistics, Springer, New York, 2005.
[37] M. Rosenblatt, Density estimates and Markov sequences, in: Nonparametric Techniques in Statistical Inference, in: Mathematical Reviews, Cambridge Univ. Press, 1970, pp. 199–210.
[38] F. Rossi, N. Villa, Support vector machine for functional data classification, Neurocomputing 69 (2006) 730–742.
[39] H. Shin, An extension of Fisher's discriminant analysis for stochastic processes, J. Multivariate Anal. 99 (2008) 1191–1216.
[40] C. Stone, Consistent non-parametric regression, Ann. Statist. 8 (1977) 348–360.
[41] T. Wagner, Strong consistency of a nonparametric estimate of a density function, IEEE Trans. Syst. Man Cybern. (1973) 289–290.