Estimating the complexity index of functional data: some asymptotics

E. G. Bongiorno[1,2,3], A. Goia[1,2,3], P. Vieu[4]

April 24, 2020

(1) Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Novara, Italy
(2) Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) - Istituto Nazionale di Alta Matematica (INdAM).
(3) Center for Artificial Intelligence - AI@UPO, Università del Piemonte Orientale
(4) Institut de Mathématiques de Toulouse, Université Paul Sabatier, Toulouse, France.

Revised Personal Version of the text of the final journal article:

E.G.Bongiorno, A.Goia, P.Vieu Estimating the complexity index of functional data: Some asymptotics. Statistics & Probability Letters Volume 161, June 2020
(doi: 10.1016/j.spl.2020.108731)

It reflects changes made in the peer review process.

Abstract

Consider a random curve valued in a general semi-metric space whose small-ball probability factorizes isolating the spatial and the volumetric term. Assume that the latter is specified and interpret its parameters as complexity indexes. An index estimate is constructed by comparing nonparametric versus parametric estimates of the volumetric factor, and various asymptotics (including weak convergence and asymptotic normality) are stated by means of U-statistics tools. As a by-product, new asymptotic results are stated for surrogate density estimation.

Keywords: Functional sample, Complexity index, Small Ball Probability, U-Statistics.

1 Introduction

Given a sample of discretized trajectories of a random function, an interesting issue, as well from theoretical as from practical point of view, is to consider the complexity nature of the underlying probability measure. In the literature various approaches have been proposed: they are usually based on measuring some (fractal) dimensions of the process, such as correlation dimension or Hausdorff one: see e.g. [2], [8], [15], and recently, [4], [5] and [6]. All these techniques are in some way related to the concept of small ball probability which can be described as follows. Given a random element X valued in a suitable semimetric space $\mathcal{F}$ and denoting by $B(\chi, h)$ the ball centered at $\chi \in \mathcal{F}$ with radius h, the small ball probability of X is $\mathbb{P}(X \in B(\chi, h))$ when h tends to zero. Under suitable conditions (see, for instance, [3], [19]), it is possible to assume that

$$\mathbb{P}(X \in B(\chi, h)) \sim \psi(\chi)\, \phi(h) \qquad \text{as } h \to 0. \tag{1}$$

The real valued function $\psi(\chi)$ plays the role of a surrogate density of the functional random element X, while the volumetric term $\phi(h)$ is independent of χ and has to be interpreted as a measure of the complexity for the process X (see [5]). To ensure identifiability of the decomposition, a normalization restriction is necessary, such as for instance

$$\mathbb{E}[\psi(X)] = 1. \tag{2}$$

To fix the ideas, note that for some special family of processes it is possible to specify the complexity function in a parametric form by means of some complexity index $\theta \in \mathbb{R}^p$ (p being some positive integer). For example, if the process has some fractal structure (as defined by [11, Definition 13.1]) then $\phi_\theta(h) = c_\theta h^\theta$, for some constant term c_θ and $\theta > 0$. Another notable example comes from Gaussian processes (see [19]), for which $\phi_\theta(h) = C_1 h^\alpha \exp\left\{-C_2/h^\beta\right\}$ with $\theta = (\alpha, \beta) \in [0, \infty) \times (0, \infty)$ and positive constants C_1, C_2. In both examples discussed just above, the estimation of θ will provide insights about the complexity of the underlying process. To appreciate why it is interesting to estimate θ, one can think to those frameworks where modeling the functional data is useful for a predictive perspective (for instance in finance, see [4, 5]) or to those situations in which the rate of convergence of nonparametric estimator depends on the small ball probability (see asymptotyc rates of convergence in [11]).

This work is not related to some special family of process. More precisely, some estimation procedure is developed under the general assumption that the volumetric component is specified by some parametric relation being of the form

$$\phi \in \{\phi_\theta, \theta \in \Theta \subset \mathbb{R}^p\}. \tag{3}$$

In this framework, one denotes by θ_0 the true value of θ. The methodology developed for estimating the parameter θ_0, consists in using a pilot nonparametric smoother of ϕ (denoted by ϕ_n) and then to build an estimate θ_n by looking for the value of $\theta \in \Theta$ which is minimizing some measure of dissimilarity between ϕ_θ and ϕ_{θ_0}.

The paper is organized as follows. First, Section 2 is devoted to the presentation of the methodology. In this section, by following ideas illustrated in [5], we define precisely an estimator θ_n for θ by minimizing the cosine measure of dissimilarity between the target ϕ_{θ_0} and the empirical nonparametric volumetric term estimate ϕ_n proposed in [6]. Then we state a wide scope of asymptotic properties for θ_n. Precisely, a weak convergence result for θ_n is derived in Section 3. Note that, as a by-product of this result, new uniform convergence for the nonparametric volumetric estimate ϕ_n are proved, completing the literature in this field (see [6]). Moreover, by exploiting the fact that ϕ_n is a second order U-statistic (see e.g. [18]), the asymptotic normality of θ_n is obtained in Section 4. Finally, a simulation study completes the analysis providing some finite sample empirical evidences on the estimator performances; see Section 5.

2 Estimating the complexity index

Let $\mathcal{F}$ be a space endowed with a suitable semi-metric and X be an $\mathcal{F}$-valued random element whose probability measure admits the factorization (1)-(2) with a volumetric term parametrized by the complexity index $\theta_0 \in \mathbb{R}^p$ according to (3). Consider $X_1, \ldots, X_n$ i.i.d. as X and the empirical estimator of ϕ defined by

$$\phi_n(h) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \mathbb{I}_{\{X_i \in B(X_j, h)\}} \qquad h \in \mathcal{H} \tag{4}$$

with $\mathcal{H} = [h_m, h_M]$, $0 < h_m < h_M$ (h_M closes to zero) and $\phi_n(h_m) > 0$. As far as we know, the literature on this estimate is limited to the pointwise weak consistency and asymptotic normal distribution properties stated [6]. This estimate will be used as a pilot tool for estimating the complexity parameter θ_0. Therefore its study is not the main purpose of this work, but it is worth being noted that some new asymptotics for ϕ_n will be obtained later on in this paper (see Proposition 1).

Starting with the pioneer paper by [7] in the usual one-dimensional linear regression framework, there is a long tradition in Nonparametric Statistics for using nonparametric smoother as pilot tool for fitting some parametric model. Ideas in [7] have been extensively applied for a wide scope of finite-dimensional problems but, as far as we know, they have been developed in infinite dimensional setting only for the standard functional linear regression purpose (see [9]). To adapt these ideas to surrogate density estimation as it is the aim of our work, one needs to define some suitable measure of dissimilarity. This is done by introducing the following cosine-dissimilarity measure:

$$\Delta(\phi_n, \phi_\theta) = 1 - \frac{\langle g(\phi_\theta), g(\phi_n)\rangle^2}{\|g(\phi_\theta)\|^2 \|g(\phi_n)\|^2}, \tag{5}$$

where $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ denote the usual inner product and the associated norm respectively in $\mathcal{L}^2_{\mathcal{H}}$, and g is a suitable continuous real-valued function defined on $\mathbb{R}$. The fact that ϕ_θ and ϕ_n are both bounded away from zero on $\mathcal{H}$ guarantee that ϕ_θ, ϕ_n, $g(\phi_\theta)$ and $g(\phi_n)$ are in $\mathcal{L}^2_{\mathcal{H}}$ and yield the wellposedness of (5). In practice, possible working choice of g are, as an instance, the identity function in the fractal case, and the logarithm for many Gaussian processes (see [5]).
Hence, an estimator θ_n of θ_0 is the minimizer of Δ:

$$\theta_n = \arg\min_{\theta \in \Theta} \Delta(\phi_n, \phi_\theta), \tag{6}$$

where Θ is a suitable compact subset of $\mathbb{R}^p$ that is supposed to contain θ_0. In the next two sections we will state asymptotic properties for this estimate.
It is worth to be mentioned that the cosine-dissimilarity is not the only possible choice as dissimilarity measure, but changing it could require different approaches and technicalities in the proofs.

3 Weak consistency

The next Theorem 1 establishes, under mild conditions the convergence in probability of θ_n to θ_0.

Theorem 1 *Assume that the model defined by (1)-(3) holds. Assume that g is Hölder continuous (i.e. $\exists C < \infty, \exists \beta > 0$, such that $\forall y_1, y_2 \in \mathbb{R}, \|g(y_1) - g(y_2)\| \leq C\|y_1 - y_2\|^\beta$) and that, for each $\theta \in \Theta$ (a compact subset of $\mathbb{R}^p$), the function ϕ_θ is continuous and increasing on $\mathcal{H}$. Then $\theta_n \longrightarrow \theta_0$* in probability, as n *diverges.*

The proof of Theorem 1 requires the following intermediary result which extends, as by-product, the existing literature on surrogate density estimation by providing a weak uniform convergence result for the empirical estimator (4) of the complexity function.

Proposition 1 *Assume that the model defined by (1)-(3) holds. If ϕ_θ is continuous and increasing on $\mathcal{H}$, then ϕ_n converges in probability to ϕ_{θ_0} uniformly on $\mathcal{H}$ as $n \to +\infty$.*

Note that the results in Theorem 1 and Proposition 1 are presented for the simple case of the empirical estimate defined in (4), but it is worth being mentioned that they still hold whenever one uses any nonparametric pointwise consistent estimator for ϕ_θ which is increasing on $\mathcal{H}$. This could concern, for instance, kernel-type estimators or kNN-type estimators as introduced in [5] and [10] (see also [20] for a general survey on nonparametric functional data analysis).

Proof of Theorem 1 Define $\delta(\theta) = \Delta(\phi_{\theta_0}, \phi_\theta)$ and $\delta_n(\theta) = \Delta(\phi_n, \phi_\theta)$ which are continuous over Θ due to the continuity of ϕ_θ. Consider

$$|\delta(\theta) - \delta_n(\theta)| = \left| \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle^2 - \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle^2 \right|$$

where $\widetilde{g}(\phi)$ is the normalized version of $g(\phi)$. Noticing that

$$\begin{aligned} \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle^2 - \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle^2 &= \left(\langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle + \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle \right) \\ &\quad \left(\langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle - \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle \right), \end{aligned}$$

by Chauchy-Schwarz inequality it follows

$$\begin{aligned} \left| \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle^2 - \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle^2 \right| &\leq \left(\|\widetilde{g}(\phi_\theta)\| \, \|\widetilde{g}(\phi_{\theta_0})\| + \|\widetilde{g}(\phi_\theta)\| \, \|\widetilde{g}(\phi_n)\| \right) \cdot \\ &\quad \cdot \left| \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) \rangle - \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_n) \rangle \right| \\ &= 2 \langle \widetilde{g}(\phi_\theta), \widetilde{g}(\phi_{\theta_0}) - \widetilde{g}(\phi_n) \rangle \leq 2 \|\widetilde{g}(\phi_{\theta_0}) - \widetilde{g}(\phi_n)\| \\ &= 2 \frac{\| \, g(\phi_{\theta_0}) \, \|g(\phi_n)\| - g(\phi_n) \, \|g(\phi_{\theta_0})\| \|}{\|g(\phi_n)\| \, \|g(\phi_{\theta_0})\|} \\ &= 2 \|\widetilde{g}(\phi_{\theta_0}) - \widetilde{g}(\phi_n)\| \end{aligned}$$

The Hölder-continuity of the function g allows to write

$$\left\|\widetilde{g}\left(\phi_{\theta_0}\right)-\widetilde{g}\left(\phi_n\right)\right\| \leq C\left\|\phi_{\theta_0}-\phi_n\right\|^\beta$$

and hence

$$\sup_{\theta \in \Theta}\left|\delta(\theta)-\delta_n(\theta)\right| \leq C\left\|\phi_{\theta_0}-\phi_n\right\|^\beta .$$

Because

$$\left\|\phi_{\theta_0}-\phi_n\right\|^{2\beta} \leq \sup_{h \in \mathcal{H}}\left(\phi_n(h)-\phi_{\theta_0}(h)\right)^{2\beta}|\mathcal{H}|^\beta \tag{7}$$

where $|\mathcal{H}|$ is the length of $\mathcal{H}$, and ϕ_n converges in probability to ϕ_{θ_0} uniformly on $\mathcal{H}$ when $n \rightarrow \infty$ (see Proposition 1), then

$$\sup_{\theta \in \Theta}\left|\delta(\theta)-\delta_n(\theta)\right| \longrightarrow 0 \qquad \text{in probability as } n \rightarrow +\infty.$$

Now, as δ is a continuous function defined over the compact set Θ, it is uniformly continuous on Θ. Moreover, thanks to the uniqueness of the minimum θ_0, it can be proved that δ satisfies the following property: for any $\varepsilon>0$, there exists $\zeta>0$ such that, for any θ, $\left\|\theta_0-\theta\right\|_p \geq \varepsilon$ implies $\left|\delta\left(\theta_0\right)-\delta(\theta)\right| \geq \zeta$ ($\|\cdot\|_p$ denotes the euclidean norm in $\mathbb{R}^p$). To show this, suppose that such assertion is false, then there would exist an $\varepsilon>0$ and a sequence $\{t_n\}$ such that, as n goes to infinity,

$$\delta\left(t_n\right) \longrightarrow \delta\left(\theta_0\right) \qquad \text{and} \qquad \left\|\theta_0-t_n\right\|_p \geq \varepsilon.$$

The latter and the fact that Θ is compact imply that t_n converges to $\theta \in \Theta \setminus \{\theta_0\}$ with $\delta(\theta)=\delta\left(\theta_0\right)$. This is against the uniqueness of the minimum of δ in Θ.
Then to prove that $\theta_n \rightarrow \theta_0$ in probability, it suffices to prove that $\delta\left(\theta_n\right) \rightarrow \delta\left(\theta_0\right)$ in probability as $n \rightarrow +\infty$. Since

$$\left|\delta\left(\theta_n\right)-\delta\left(\theta_0\right)\right| \leq\left|\delta\left(\theta_n\right)-\delta_n\left(\theta_n\right)\right|+\left|\delta_n\left(\theta_n\right)-\delta\left(\theta_0\right)\right|$$

where

$$\left|\delta\left(\theta_n\right)-\delta_n\left(\theta_n\right)\right| \leq \sup_{\theta \in \Theta}\left|\delta(\theta)-\delta_n(\theta)\right|$$

and

$$\left|\delta_n\left(\theta_n\right)-\delta\left(\theta_0\right)\right| \leq\left|\sup_{\theta \in \Theta} \delta_n(\theta)-\sup_{\theta \in \Theta} \delta(\theta)\right| \leq \sup_{\theta \in \Theta}\left|\delta_n(\theta)-\delta(\theta)\right|,$$

it holds

$$\left|\delta\left(\theta_n\right)-\delta\left(\theta_0\right)\right| \leq 2 \sup_{\theta \in \Theta}\left|\delta(\theta)-\delta_n(\theta)\right| \longrightarrow 0 \qquad \text{in probability as } n \rightarrow +\infty$$

that concludes the proof.

Proof of Proposition 1 Because ϕ_{θ_0} is increasing and continuous on $\mathcal{H}$, for any $\varepsilon>0$, it is possible to consider $h_m=t_0<t_1<\ldots<t_N=h_M$ so that

$$\phi_{\theta_0}\left(t_j\right)-\phi_{\theta_0}\left(t_{j-1}\right)<\varepsilon \qquad j=1, \ldots, N. \tag{8}$$

Thus

$$\mathbb{P}\left(\sup_{h \in \mathcal{H}}\left|\phi_n(h)-\phi_{\theta_0}(h)\right|>\varepsilon\right)=\mathbb{P}\left(\max_{j=1, \ldots, N} \sup_{t_{j-1}<h \leq t_j}\left|\phi_n(h)-\phi_{\theta_0}(h)\right|>\varepsilon\right)$$

$$\leq \mathbb{P}\left(\max_{j=1, \ldots, N}\left|\phi_n\left(t_j\right)-\phi_{\theta_0}\left(t_{j-1}\right)\right|>\varepsilon\right)=\mathbb{P}\left(\max_{j=1, \ldots, N}\left|\phi_n\left(t_j\right)-\phi_{\theta_0}\left(t_j\right)+\phi_{\theta_0}\left(t_j\right)-\phi_{\theta_0}\left(t_{j-1}\right)\right|>\varepsilon\right)$$

$$\leq \mathbb{P}\left(\max_{j=1, \ldots, N}\left|\phi_n\left(t_j\right)-\phi_{\theta_0}\left(t_j\right)\right|>\varepsilon\right)+\mathbb{P}\left(\max_{j=1, \ldots, N}\left|\phi_{\theta_0}\left(t_j\right)-\phi_{\theta_0}\left(t_{j-1}\right)\right|>\varepsilon\right)$$

In latter, the second addend is null thanks to (8), whereas, for the first one it holds

$$\mathbb{P}\left(\max_{j=1,\ldots,N}\left|\phi_n\left(t_j\right)-\phi_{\theta_0}\left(t_j\right)\right|>\varepsilon\right)\leq\max_{j=1,\ldots,N}\mathbb{P}\left(\left|\phi_n\left(t_j\right)-\phi_{\theta_0}\left(t_j\right)\right|>\varepsilon\right).$$

Thanks to the pointwise consistency of ϕ_n to ϕ_{θ_0}, the postulated result follows immediately.

4 Asymptotic normality

In this section, a central limit theorem for θ_n is proved. One of the key arguments exploited in the proof is that ϕ_n is a U-statistics. In what follows, $\delta\left(\theta\right)=\Delta\left(\phi_{\theta_0},\phi_\theta\right)$ and $\delta_n\left(\theta\right)=\Delta\left(\phi_n,\phi_\theta\right)$ as already introduced at the beginning of the proof of Theorem 1.

Theorem 2 *Under the same hypothesis of Theorem 1 and assuming that g is $\mathcal{C}^2\left(0,+\infty\right)$ (i.e. twice derivable with continuous derivatives over $\left(0,+\infty\right)$) with non-null first derivative, δ is $\mathcal{C}^2\left(\Theta\right)$ and strictly convex over Θ, then as $n\to+\infty$*

$$\sqrt{n}\frac{\left\|g\left(\phi_{\theta_0}\right)\right\|^4}{2}\nabla^2\delta\left(\theta_0\right)\left(\theta_n-\theta_0\right)\sim\mathcal{N}\left(0,\Gamma\right)$$

where $\nabla^2\delta$ is the Hessian matrix of δ, and Γ is given in (11).

Proof of Theorem 2. In the treated framework, one has $\delta\left(\theta\right)=1-\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\widetilde{g}\left(\phi_\theta\right)\right\rangle^2$ with $\theta\in\Theta$ and $\widetilde{g}\left(\phi\right)$ being the normalized version of $g\left(\phi\right)$.
Moreover denoting by t_r the r-th element of θ, the gradient and the Hessian matrix of $\delta\left(\theta\right)$ with respect to θ write:

$$\nabla\delta\left(\theta\right)=-2\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\widetilde{g}\left(\phi_\theta\right)\right\rangle\left[\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\frac{\partial}{\partial t_r}\widetilde{g}\left(\phi_\theta\right)\right\rangle\right]_{r=1,\ldots,p}$$

$$\nabla^2\delta\left(\theta\right)=-2\left[\begin{array}{c}\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\frac{\partial}{\partial t_s}\widetilde{g}\left(\phi_\theta\right)\right\rangle\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\frac{\partial}{\partial t_r}\widetilde{g}\left(\phi_\theta\right)\right\rangle+\\+\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\widetilde{g}\left(\phi_\theta\right)\right\rangle\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\frac{\partial^2}{\partial t_r\partial t_s}\widetilde{g}\left(\phi_\theta\right)\right\rangle\end{array}\right]_{r,s=1,\ldots,p}$$

and

$$\nabla\delta\left(\theta_0\right)=\mathbf{0},\qquad\nabla^2\delta\left(\theta_0\right)=-2\left[\left\langle\widetilde{g}\left(\phi_{\theta_0}\right),\frac{\partial^2}{\partial t_r\partial t_s}\widetilde{g}\left(\phi_{\theta_0}\right)\right\rangle\right]_{r,s=1,\ldots,p}.\tag{9}$$

Consider $\delta_n\left(\theta\right)=1-\left\langle\widetilde{g}\left(\phi_n\right),\widetilde{g}\left(\phi_\theta\right)\right\rangle^2$ and the Taylor expansion of $\nabla\delta_n\left(\theta\right)$ about θ_0 with Lagrange remainder:

$$\nabla\delta_n\left(\theta\right)=\nabla\delta_n\left(\theta_0\right)+\nabla^2\delta_n\left(\theta^\star\right)\left(\theta-\theta_0\right)$$

where the elements $t_r^\star$ of $\theta^\star\in\Theta$ stay between t_r and $t_{0,r}$ (the r-th element of θ_0). Evaluate it at its minimum θ_n, so that $\nabla\delta_n\left(\theta_n\right)=\mathbf{0}$, and since $\nabla^2\delta\left(\theta\right)$ is invertible, direct calculations lead to

$$\theta_n-\theta_0=-\left(\nabla^2\delta_n\left(\theta^\star\right)\right)^{-1}\nabla\delta_n\left(\theta_0\right)\tag{10}$$

where

$$\nabla\delta_n\left(\theta_0\right)=-2\left\langle\widetilde{g}\left(\phi_n\right),\widetilde{g}\left(\phi_{\theta_0}\right)\right\rangle\left[\left\langle\widetilde{g}\left(\phi_n\right),\frac{\partial}{\partial t_r}\widetilde{g}\left(\phi_{\theta_0}\right)\right\rangle\right]_{r=1,\ldots,p}$$

and

$$\nabla^2\delta_n\left(\theta^\star\right)=-2\left[\begin{array}{c}\left\langle\widetilde{g}\left(\phi_n\right),\frac{\partial}{\partial t_s}\widetilde{g}\left(\phi_{\theta^\star}\right)\right\rangle\left\langle\widetilde{g}\left(\phi_n\right),\frac{\partial}{\partial t_r}\widetilde{g}\left(\phi_{\theta^\star}\right)\right\rangle+\\+\left\langle\widetilde{g}\left(\phi_n\right),\widetilde{g}\left(\phi_{\theta^\star}\right)\right\rangle\left\langle\widetilde{g}\left(\phi_n\right),\frac{\partial^2}{\partial t_r\partial t_s}\widetilde{g}\left(\phi_{\theta^\star}\right)\right\rangle\end{array}\right]_{r,s=1,\ldots,p}.$$

Thanks to Proposition 1 and Equation (9), one gets that $\langle \widetilde{g}(\phi_n), \widetilde{g}(\phi_{\theta_0})\rangle$ and $\left[\left\langle \widetilde{g}(\phi_n), \frac{\partial}{\partial t_r}\widetilde{g}(\phi_{\theta_0})\right\rangle\right]_{r=1,\dots,p}$ converge to one and the null vector respectively, in probability as $n \to +\infty$. Moreover, by Theorem 1, the definition of $\theta^\star$ and thanks to Proposition 1, $\left\langle \widetilde{g}(\phi_n), \frac{\partial}{\partial t_r}\widetilde{g}(\phi_{\theta^\star})\right\rangle$ goes to zero in probability as n diverges, for any r, while $\nabla^2\delta_n(\theta^\star)$ tends to $\nabla^2\delta(\theta_0)$. Summarizing,

$$\theta_n - \theta_0 \longrightarrow 2\nabla^2\delta(\theta_0)^{-1}\left[\left\langle \widetilde{g}(\phi_n), \frac{\partial}{\partial t_r}\widetilde{g}(\phi_{\theta_0})\right\rangle\right]_{r=1,\dots,p}$$

in probability as $n \to +\infty$. Hence, to derive the asymptotic distribution of $\theta_n - \theta_0$, thanks to the Slutsky Theorem, it is sufficient to study the law of the random vector:

$$\left[\left\langle \widetilde{g}(\phi_n), \frac{\partial}{\partial t_r}\widetilde{g}(\phi_{\theta_0})\right\rangle\right]_{r=1,\dots,p}$$

that equals

$$\left[\frac{\langle g(\phi_n), \eta_{0,r}\rangle}{\|g(\phi_n)\|\,\|g(\phi_{\theta_0})\|^3}\right]_{r=1,\dots,p}$$

where

$$\eta_{0,r} = g'(\phi_{\theta_0})\left(\frac{\partial\phi_{\theta_0}}{\partial t_r}\right)\|g(\phi_{\theta_0})\|^2 - g(\phi_{\theta_0})\int g(\phi_{\theta_0})\,g'(\phi_{\theta_0})\left(\frac{\partial\phi_{\theta_0}}{\partial t_r}\right) \qquad r = 1,\dots,p.$$

As a consequence of Proposition 1, as n diverges, $\|g(\phi_n)\|$ tends to the constant $\|g(\phi_{\theta_0})\|$ in probability. Hence, evoking again the Slutsky Theorem it remains to study the asymptotic distribution of the random vector whose entries are

$$\langle g(\phi_n), \eta_{0,r}\rangle = \sqrt{n}\,\langle g(\phi_n) - g(\phi_{\theta_0}), \eta_{0,r}\rangle \qquad r = 1,\dots,p,$$

because $\langle g(\phi_{\theta_0}), \eta_{0,r}\rangle = 0$ for any $r = 1,\dots,p$.
For any $h \in \mathcal{H}$, the Taylor expansion of g at $\phi_{\theta_0}(h)$ provides

$$\sqrt{n}\,(g(\phi_n(h)) - g(\phi_{\theta_0}(h))) = \sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))\,g'(\phi_{\theta_0}(h)) + o\left(\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))\right).$$

The fact that $\phi_n(h)$ is a U-Statistic guarantees that the finite dimensional distributions of $\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))$ converge to those of a centered Gaussian process with covariance function (see [13] and [22])

$$\varsigma(h, h') = \sum_{i,j}\sum_{k,m}\mathbb{P}\left(\{\|X_i - X_j\| \le h\} \cap \{\|X_k - X_m\| \le h'\}\right) - 4\phi_{\theta_0}(h)\,\phi_{\theta_0}(h').$$

As a consequence

$$\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))\,g'(\phi_{\theta_0}(h))$$

converges to a centered Gaussian process with covariance function

$$\sigma(h, h') = g'(\phi_{\theta_0}(h))\,\varsigma(h, h')\,g'(\phi_{\theta_0}(h'))$$

and $\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))$ is bounded in probability, so that:

$$o\left(\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))\right) = o(1) \qquad h \in \mathcal{H}.$$

Hence $[\sqrt{n}\,\langle g(\phi_n), \eta_{0,r}\rangle]_{r=1,\dots,p}$ is asymptotically distributed as a centered Gaussian random vector with variance

$$\Gamma = [\langle \eta_{0,s}, \Sigma[\eta_{0,r}]\rangle]_{r,s=1,\dots,p} \qquad (11)$$

where Σ is the covariance operator associated to the covariance function σ.

5 Simulation

In this section the results of a brief simulation study are reported in order to provide some empirical evidences on the theoretical results derived in the previous sections. In practice, the empirical distributions of the estimator are obtained starting from 1000 Monte Carlo samples of independent random curves $X_1, \ldots, X_n$ generated according to a random process X and discretized over a mesh of 100 equispaced points, with $n = 50, 100, 200, 500$.
We used both the noised bi–dimensional process

$$X(t) = \sqrt{2}\left(A \sin(\pi t) + B \cos(\pi t)\right) + \sigma \mathcal{E}(t) \qquad t \in [0,1]$$

(where, A and B are independent $\mathcal{N}(0,1)$, $\mathcal{E}(t)$ is a standard Gaussian white noise and $\sigma = 0.02$) and a Brownian Bridge on $[0,1]$. We recall that, when h tends to zero, $\phi(h)$ is proportional to h^{θ} in the first case whereas, in the second case, $\log(\phi(h))$ is proportional to $h^{-\theta}$, with $\theta = 2$ in both cases.
To make calculations, the interval $\mathcal{H}$ has been discretized in a grid of approximately $\sqrt{2n}$ equispaced points between $h_m = 1/(10n)$ and $h_M = 3/n$, so that the range of $\mathcal{H}$ became closer to zero as the sample size increases.
For each used process and sample size, some synthetic indicators (mean, standard deviation and median) of the generated distributions were calulated and the Shapiro-Wilk test of normality was performed. The results are collected in Table 1. Reading them, it emerges that in the finite dimensional case no evident bias are present whereas in the infite dimensional one there is a bias which decreases with n and that could be linked with the choice of $\mathcal{H}$. The variability of the estimator decreases, as is to be expected, with n. The p-value associated to the normality test increases with n providing an evidence in favour of the normality of the distribution of the estimators, at least for moderately large sample sizes.
More numerical studies can be found in [5].

Process	n	Mean	St.Dev.	Median	p–value
Finite dimensional	50	2.003	0.248	1.995	0.0013
(with $\theta = 2$)	100	1.999	0.178	1.993	0.3515
	200	1.998	0.126	1.995	0.4573
	500	2.004	0.081	2.004	0.3566
Brownian Brigde	50	2.205	0.318	2.193	0.0000
(with $\theta = 2$)	100	2.184	0.224	2.179	0.0009
	200	2.137	0.159	2.135	0.1579
	500	2.101	0.103	2.101	0.1662

Table 1: Some statistics on the distributions of the estimated complexity index under different experimental conditions.

6 Concludings

This paper provides theoretical insights for estimating the complexity of some functional random element that enrich the methodological framework described and implemented in [4, 5] opening the route for deeper descriptive analysis of Functional Data. In this sense, the paper aims to contribute to the broader scientific debate on Functional Statistics. To have an idea on the wide scope of methodologies introduced along the recent years in such discipline, one can consult the monographes [11], [14], [17], [21] or the special issues [1], [12] and [16].

Acknowledgements

E. G. Bongiorno and A. Goia thank Università del Piemonte Orientale for the financial support.

References

[1] Aneiros, G., Cao, R., Fraiman, R. and Vieu, P. (2019). Editorial for the Special Issue on Functional Data Analysis and Related Topics. Journal of Multivariate Analysis, 170, 1–2.

[2] Bardet J.-M. (1997), Tests d'autosimilarité des processus gaussiens. Dimension fractale et dimension de corrélation, Thèse, Université Paris-Sud, France.

[3] Bongiorno, E.G., Goia, A. (2017). Some insights about the Small Ball Probability factorization for Hilbert random elements. Statistica Sinica, 27, 1949–1965.

[4] Bongiorno E.G., Goia A., Vieu P. (2017). On the Geometric Brownian Motion assumption for financial time series. In: Functional Statistics and Related Fields, Aneiros G., Bongiorno E.G., Cao R., Vieu P. (Eds). Springer.

[5] Bongiorno, E.G., Goia, A., Vieu, P. (2018). Evaluating the complexity of some families of functional data. SORT – Statistics and Operations Research Transactions, 42, 27–44.

[6] Bongiorno, E. G., Goia, A., Vieu, P. (2019). Modeling functional data: a test procedure. Computational Statistics, 34(2), 451-468.

[7] Cristóbal Cristóbal, J. A., Faraldo Roca, P. and Gonzàlez Manteiga, W. (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. Annals of Statistics, 15(2), 603–609.

[8] Cutler, C. D., Dawson, D. A. (1990). Nearest-neighbor analysis of a family of fractal distributions. The Annals of Probability, 256-271.

[9] Delsol, L. (2013). No effect tests in regression on functional variable and some applications to spectrometric studies. Comput. Statist. 28(4), 1775–1811.

[10] Ferraty, F., Kudraszow, N. and Vieu, P. (2012). Nonparametric estimation of a surrogate density function in infinite-dimensional spaces. Journal of Nonparametric Statistics, 24, 447–464.

[11] Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis. Springer Series in Statistics. Springer, New York.

[12] Goia, A., Vieu, P. (2016). Editorial – An introduction to recent advances in high/infinite dimensional Statistics. Journal of Multivariate Analysis, 146, 1–6.

[13] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Annals of Mathematical Statistics, 19, 293–325 .

[14] Horvath, L. and Kokoszka, P. (2012). Inference for Functional Data with Applications. Springer Series in Statistics. Springer, New York.

[15] Kawaguchi, A. (2002). Estimating the correlation dimension from chaotic dynamical systems by U-statistics. Bulletin of Informatics and Cybernetics, 34(2), 143–150.

[16] Kokoszka, P., Oja, H., Park, B., Sangalli, L. (2017). Special issue on functional data analysis. Econometrics and statistics, 1(C), 99–100.

[17] Kokoszka, P., Reimherr, M. (2017). Introduction to Functional Data Analysis. Chapman and Hall.

[18] Lee, J. (1990). U-statistics: Theory and Practice. Routledge.

[19] Li, W. V. and Shao, Q.-M. (2001). Gaussian processes: inequalities, small ball probabilities and applications. In Stochastic Processes: Theory and Methods, vol. 19 of Handbook of Statist. North-Holland, Amsterdam, 533–597.

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

[21] Ramsay, J.O. and Silverman, B.W. (2005). Functional Data Analysis, 2nd Edition. Springer Series in Statistics. Springer, New York.

[22] Silverman, B. W. (1983). Convergence of a class of empirical distribution functions of dependent random variables. The Annals of Probability, 11(3), 745–751.