MOX–Report No. 39/2014

Statistical inference for functional data based on a generalization of Mahalanobis distance

Ghiglietti, A.; Paganoni, A.M.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Statistical inference for functional data based on a generalization of Mahalanobis distance

Andrea Ghiglietti[a], *Anna Maria Paganoni*[b]

[a] Department of Mathematics "F. Enriques"
Università degli Studi di Milano
via Saldini 50, 20133 Milan, Italy
andrea.ghiglietti@unimi.it
[b] Department of Mathematics "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milan, Italy
anna.paganoni@polimi.it

Abstract

In this paper we propose a generalization of Mahalanobis distance that extends the usual multivariate one to functional data generated by stochastic processes. We show that this distance is well defined in $L^2(T)$ and achieves both the goals of (i) considering all the infinite components of data basis expansion and (ii) keeping the same ideas on which is based the Mahalanobis distance. This new mathematical tool is adopted in an inferential context to construct tests on the mean of Gaussian processes for one and two populations. The tests are constructed assuming the covariance structure to be either know or unknown. The power of all the critical regions has been computed analytically. A wide discussion on the behavior of these tests in terms of their power functions is realized, supported by some simulation studies.

Keywords:Functional Data, Distances in L^2, Gaussian Processes, Inference on the mean.
AMS Subject Classification:62H15, 62M86.

1 Introduction

Nowadays, many scientific fields as biostatistics, econometrics, etc.., deal with data concerning continuous phenomenons of time or space. For this reason, more and more often observations are modeled as a sample of i.i.d. random functions $X_1, .., X_n$, typically in $L^2(T)$ with T compact set of $\mathbb{R}$. The related mean function m and the covariance operator V usually represent the main object of the statistical investigation. Functional Data Analysis (FDA) gathers all the statistical models that consider data as functional objects and it is remarkable the increase of its importance in the statistical literature during the last decade. A complete overview on the most common statistical methods, computational details and case studies in FDA, is provided, for example, in the books by Ramsey and Silverman (2002),

Ramsey and Silverman (2005) and Ferraty and Vieu (2006). As highlighted in this literature, a central role in this context is carried out by the Functional Principal Component Analysis (FPCA). The basic tool for this analysis is the Karhunen-Loève (KL) expansion, that decomposes a random function $X(t)$ in a sum of the mean $m(t)$ and a series of orthonormal functions $\theta_k(t)$, each one multiplied by zero-mean uncorrelated random variables $\sqrt{\lambda_k}Z_k$, ($\lambda_k > 0$, $\mathbb{V}ar(Z_k) = 1$). The othonormal basis $(\theta_k)_k$ is composed by the eigenfunctions of the covariance operator V of X, while the coefficient variances $(\lambda_k)_k$ are its eigenvalues. Hence, the dynamic of the random function $X(t) - m(t)$ can be fully described by the eigenstructure of V and the distribution of the sequence $(Z_k)_k$. Since this paper is focused on Gaussian processes, in our case $(Z_1, Z_2, ..)$ are i.i.d. standard normal variables. As discussed in Benko *et al.* (2009), the analysis of the principal components seems to be more important in the functional context than in the multivariate framework. In fact, it is one of the most feasible way to reduce data dimensionality. Moreover, in FPCA the principal components are interpreted as the modes of variation of $X(t)$ along t, which is much more interpretable than arguments concerning the correlation of principal components with original variables adopted in multivariate PCA. When the goal of the analysis is to describe the shape of $X(t)$, the first K principal components θ_k, $k = 1, .., K$, usually contain all the information needed to represent data. Nevertheless, when to goal is to make inference on infinite dimensional objects, as the mean function $m(t)$, considering a fixed number of components may lead to lose some information on the distribution of $X(t)$ and to provide meaningless results.

Despite of the great interest in the FPCA, many inferential procedures adopted in the multivariate PCA have not been extended yet to the functional case. For instance, in the multivariate context the inference on the mean is typically based on the Mahalanobis distance, since it is the best way to measure the distance between elements because it takes into account the correlation among the variables and their variability. However, when data belongs to an infinite dimensional space, as $L^2(T)$, the Mahalanobis distance is in general undefined and the inference is usually realized by considering only the first K principal components. This approach is pretty far from the Mahalanobis idea, which weights the distances along all the components with the inverse of their variability. In this paper we propose a generalization of Mahalanobis distance that extends the usual multivariate one to functional data generated by stochastic processes. The new metrics has been obtained after noticing a quite unconventional way to derive the classical Mahalanobis distance. We show that this distance is well defined in $L^2(T)$ and achieves both the goals of (i) considering all the infinite components of data basis expansion and (ii) keeping the same ideas on which is based the Mahalanobis distance (see Section 2). This new mathematical tool is adopted in an inferential context to construct tests on the mean of Gaussian processes for one and two populations. The tests are constructed assuming the covariance structure to be either know, (3.2) and (3.4), or unknown, (4.8) and (4.9). An extension of inferential procedures based on the Mahalanobis distance has been proposed in Secchi *et al.* (2013), by considering multivariate Gaussian data with a number of variables increasing to infinity. Nevertheless, the test proposed in Secchi *et al.* (2013) doesn't consider directly functional data and it is not applicable if the covariance is assumed to be known. Moreover, the test proposed in Secchi *et al.* (2013) doesn't consider the differences among functions along the components unexplored by the data.

In a typical FDA framework, the independent realizations of X are not observed

directly as curves, but we can only measure them at some discrete values, often with an independent additive error term. Then, the first fundamental step for the inference is usually an estimation of the mean function $m(t)$ via smoothing techniques, see for instance Ramsey and Silverman (2002), Ruppert *et al.* (2003), Ramsey and Silverman (2005), Cuevas *et al.* (2006), Ferraty and Vieu (2006), Li and Hsing (2010), Degras (2011). In the literature, we find methods based on splines, which can be penalized as in Ramsey and Silverman (2005) or free-knot as in Gervini (2006), methods which deal with ridge-type least squares estimates as in Rice and Silverman (1991), or kernel smoothers as in Yao (2007), Zhang and Chen (2007) and Benko *et al.* (2009). After estimation, the inference is usually computed by constructing confidence bands for $m(t)$; see for instance see Yao *et al.* (2005) and Ma *et al.* (2012) for sparse longitudinal data, Bunea *et al.* (2011), Degras (2011) and Cao *et al.* (2012) for dense functional data. Since the main purpose of this paper is focused on the inference of the mean function, we assume to observe perfectly the random curves over the whole range without any noise effect, as already considered in many other works, see Luschgy (1991), Bosq (2000), Ferraty and Vieu (2006). In Luschgy (1991), our framework is presented as a typical signal-plus-noise model with deterministic signal $m(t)$ and noise $X(t) - m(t)$. In that paper, a locally most powerful test for smooth function $m(t)$ is derived and its asymptotic properties are discussed. However, in Luschgy (1991) the mean $m(t)$ is forced to belong to a known class of functions indexed by a real parameter, and the resulting inference is only on that parameter, while here we let the mean function to be any element in $L^2(T)$.

The inference of the mean of a Gaussian process has been treated also in Bunea *et al.* (2011), with observations at discrete times and additive noise terms. They propose a computationally simple method based on threshold projection estimators, providing fully data-driven estimator and confidence sets for $m(t)$. However, the confidence level is achieved only asymptotically and for any sample size n the balls are constructed with a fixed number of components, while the inferential procedures proposed in this paper consider all the infinite dimensional space. In fact, instead of the correction implemented to construct the threshold in Bunea *et al.* (2011), we adopt a distance d_p tuned by a parameter that smoothly determines how to weight the contributions along all the infinite components of $L^2(T)$. Moreover, this parameter plays a crucial role in the power of the tests (see Section 5), which is a topic rarely discussed in literature. In fact, inferential tools in functional data analysis are typically based on confidence bands and so the properties of the tests under the alternative hypothesis are barely considered. In this paper, not only the powers of all the critical regions are computed analytically, but Section 5 is mainly dedicated to discuss this point. The confidence band suggested in Cao *et al.* (2012) takes into account the variability of the random process X along the interval T, but the effect of the covariance among different times is included in the model only through a parameter. Then, it's easy to visualize and to interpret the confidence bands, but the level is reached only asymptotically. The spline estimator and the confidence band for the mean function proposed in Cao *et al.* (2012) are asymptotically the same as if all random trajectories are observed entirely and without errors. This result is achieved by letting both the sample size and the number of observation points opportunely increase to infinity in order to control the covariance of the mean estimator. As mentioned in Cao *et al.* (2012), this provides a theoretical justification for treating functional data as perfectly recorded

random curves over the whole range, as done here. Moreover, in many applications it is reasonable to model also the noise as a Gaussian process $\epsilon(t)$ independent of $X(t)$ with covariance operator V_ϵ, since random errors in correspondence to closed times are usually correlated. Then, each observed random curve can be model as a Gaussian process with the same mean $m(t)$ and covariance operator $V + V_\epsilon$, and the tests proposed in this paper can be perfectly applied to this framework.

When the covariance of the process X is unknown, the elements λ_k, θ_k of the eigen-structure of V used to compute the distance d_p are estimated respectively by the eigenvalues $\widehat{\lambda}_k$ and eigenfunctions $\widehat{\theta}_k$ of the empirical covariance operator $\widehat{V}_n$ computed from data. Naturally, an eigenfunction θ_k is univocally determined (up to sign) only if corresponding eigenvalue λ_k has multiplicity one. Then, many methods in the functional principal component estimation usually require some necessary regularity condition on the unknown covariance function, see for instance Bosq (2000). However, we don't need these assumptions since the distance d_p defined in this paper is invariant with respect to both the sign of θ_k and the choice of the basis in the eigenspaces associated to λ_k with multiplicity greater than one. Starting from discrete and noisy observation, the usual approach is to estimate the sample curves nonparametrically (e.g., by B-splines) and then to derive the principal components from the empirical covariance function, see Besse and Ramsay (1986), Ramsey and Dalzell (1991). Rice and Silverman (1991), Pezzulli and Silverman (1993) or Silverman (1996) introduce smoothing techniques in the eigenanalysis and Lacantore *et al.* (1999) investigate the robustness in the principal component estimation. Yao *et al.* (2005) and Hall and Hosseini-Nasab (2006) propose nonparametric estimation methods for sparse data. In Benko *et al.* (2009) inferential procedures on testing the eigenstructure of V based on a bootstrap technique is studied. This is a generalization of the asymptotic results on the empirical eigenvalues and eigenfunctions distribution realized in Dauxois *et al.* (1982) in the case of functions directly observable. Moreover, Benko *et al.* (2009) shows that under mild conditions the error in the estimation derived from considering discrete and noisy data is first-order asymptotically negligible, and so the inference may proceed as if the functions were directly observed.

In Section 2 the new distance d_p is introduced and its mathematical properties are discussed. Then, in Section 3 and 4 we construct critical regions for the inference on the mean of Gaussian processes for one and two populations. In Section 3 we analyze the situation of known covariance function, while in Section 4 the covariance is estimated from data. For all the proposed critical regions, the analytic expression of the power is computed as a function of the real mean in $L^2(T)$. Finally, in Section 5 we discuss the statistical properties of the tests proposed in the paper, specially focusing on the power study, guided by some simulation studies.

2 A generalization of Mahalanobis distance in functional framework

In this section, we introduce a new metric defined in an infinite dimensional space, which generalizes the Mahalanobis distance used in the multivariate context. Let first consider a finite dimensional framework. Let $\underline{X} \in \mathbb{R}^K$, $K \in \mathbb{N}$, be a random element with mean $\underline{m} \in \mathbb{R}^K$ and covariance matrix $V \in \mathbb{R}^{K \times K}$, and call $\underline{\theta}_1, ..., \underline{\theta}_K$ the eigenvectors of V and $\lambda_1 \geq ... \geq \lambda_K > 0$ the associated eigenvalues. For any couple $(\underline{y}, \underline{w})$ of elements of $\mathbb{R}^K$, the usual metric is the euclidian distance

$d_e(\underline{y},\underline{w}) = \sqrt{\sum_{k=1}^{K}((\underline{y}-\underline{w})^T\underline{\theta}_k)^2}$, where $\underline{a}^T\underline{b} = \sum_{i=1}^{K} a_i b_i$ is the usual inner product in $\mathbb{R}^K$. However, when $\underline{y}$ and $\underline{w}$ are realization of $\underline{X}$, the Mahalanobis distance, i.e.

$$d_M(\underline{y},\underline{w}) = \sqrt{\sum_{k=1}^{K} \frac{\left((\underline{y}-\underline{w})^T\underline{\theta}_k\right)^2}{\lambda_k}},$$

may be more useful since it takes into account the correlations and the variability described by the covariance structure of $\underline{X}$. Here, we are going to highlight an interesting relation between these two distances. Let introduce the function $f(\cdot;\underline{y},\underline{w}) : \mathbb{R}^+ \to \mathbb{R}^+$ defined as follows

$$f(c;\underline{y},\underline{w}) = \sum_{k=1}^{K} f_k(c;\underline{y},\underline{w}) = \sum_{k=1}^{K} \left((\underline{y}-\underline{w})^T\underline{\theta}_k\right)^2 \cdot \exp(-\lambda_k c).$$

First, note that, for any fixed $c \in \mathbb{R}^+$, $\sqrt{f(c;\underline{y},\underline{w})}$ is a distance, since it can be seen as the euclidian distance among $\widetilde{\underline{y}} := \sum_{k=1}^{K}\left((\underline{y}-\underline{w})^T\underline{\theta}_k \exp(-\lambda_k c/2)\right)\cdot\underline{\theta}_k$ and $\widetilde{\underline{w}} := \sum_{k=1}^{K}\left((\underline{y}-\underline{w})^T\underline{\theta}_k \exp(-\lambda_k c/2)\right)\cdot\underline{\theta}_k$. Moreover, we have:

(1) $f(c;\underline{y},\underline{w})$ is non increasing in c;

(2) $\lim_{c\to\infty} f_k(c;\underline{y},\underline{w}) = 0$, $\forall\lambda_k > 0$;

(3) $f(0;\underline{y},\underline{w}) = d_e^2(\underline{y},\underline{w})$.

Then, for any fixed $c \in \mathbb{R}^+$, $\sqrt{f(c;\underline{y},\underline{w})}$ represents a distance among $\underline{y}$ and $\underline{w}$ that is less than or equal to the euclidian distance. In particular, $f_k(c;\underline{y},\underline{w})$ is the contribution to this distance along the component $\underline{\theta}_k$. Moreover, as c increases the term $f_k(c;\underline{y},\underline{w})$ get smaller. The decreasing rate of $f_k(c;\underline{y},\underline{w})$ is ruled by λ_k: the greater is λ_k the faster $f_k(c;\underline{y},\underline{w})$ vanishes. Then, the term $\exp(-\lambda_k c)$ is a quantity that penalizes the contribution of the euclidian distance along θ_k and this penalization is strong for high λ_k and irrelevant for low λ_k. In the Mahalanobis distance a similar behaviour is obtained by rescaling with respect to λ_k; this rescaling makes stronger the components with low λ_k and weaker the ones with high λ_k. Then, we can think to measure how fast $f_k(c;\underline{y},\underline{w})$ vanishes by integrating over c, so we obtain

$$\int_0^\infty f_k(c;\underline{y},\underline{w})dc \;=\; \frac{\left((\underline{y}-\underline{w})^T\underline{\theta}_k\right)^2}{\lambda_k}$$

So the Mahalanobis distance can be seen as the square root of the area below the function $f(c;\underline{y},\underline{w})$:

$$d_M(\underline{y},\underline{w}) = \sqrt{\sum_{k=1}^{K} \frac{\left((\underline{y}-\underline{w})^T\underline{\theta}_k\right)^2}{\lambda_k}} = \sqrt{\sum_{k=1}^{K}\int_0^\infty f_k(c;\underline{y},\underline{w})dc} = \sqrt{\int_0^\infty f(c;\underline{y},\underline{w})dc} \tag{2.1}$$

We want to extend these ideas to a functional framework. Let y and w be elements of $L^2(T)$ realization of a stochastic process $X \in L^2(T)$. Let $m(t) = \mathbb{E}[X(t)]$ be the mean function and V the covariance operator of X, i.e. V is a

linear compact integral operator from $L^2(T)$ to $L^2(T)$ acting as follows: $(Va)(s) = \int_T v(s,t)a(t)dt \ \forall a \in L^2(T)$, where v is the covariance function defined as $v = \mathbb{E}[(X(t) - m(t))(X(s) - m(s))]$. Then, denote with $(\lambda_k)_k$ and $(\theta_k)_k$ the sequences of the eigenvalues and the associated eigenfunctions of v, respectively. Let $\langle a, b\rangle = \int_T a(t)b(t)dt$ be the usual inner product in $L^2(T)$, so the Mahalanobis distance is

$$d_M(y,w) = \sqrt{\sum_{k=1}^{\infty} \frac{(\langle y - w, \theta_k\rangle)^2}{\lambda_k}},$$

which could be undefined since the series can diverge for some $y, w \in L^2(T)$. For this reason, a typical practice is to fix an integer $K \in \mathbb{N}$ and consider the truncated version of the Mahalanobis distance, summing up only the first K components. However, when this approach is used to measure the entire space $L^2(T)$, we can point out two main drawbacks:

- the contribution given by the projections in the space orthogonal to $\theta_1, .., \theta_K$ is not considered in the distance. Then, for any choice of K, we may have $y, w \in L^2(T)$ such that the truncated Mahalanobis distance is arbitrarily small and the euclidian distance is arbitrarily large, which seems unreasonable.
- all the contributions of the $L^2(T)$ distance are basically multiplied by $1/\lambda_k \cdot \mathbf{1}_{\{\lambda_k \geq \lambda_K\}}$, which is not decreasing in λ_k. This is incoherent with the idea of penalizing the $L^2(T)$ distance with a term that is inversely proportional to the size of the corresponding eigenvalue λ_k.

Our goal is to use the new representation (2.1) of the Mahalanobis distance to compute a metric that solves these problems. Note that (2.1) can be straightforwardly extended to the functional framework. In fact, for any couple of elements of $L^2(T)$ (y, w) we can define the function $f(\cdot; y, w)$ as

$$f(c; y, w) = \sum_{k=1}^{\infty} (\langle y - w, \theta_k\rangle)^2 \cdot \exp(-\lambda_k c).$$

Notice that this series is finite for any $c \in \mathbb{R}^+$. As explained before, the Mahalanobis distance can be computed as $\sqrt{\int_0^\infty f(c; y, w)dc}$. However, in the infinite dimensional case, when the Mahalanobis distance is not defined, the function $f(c; y, w)$ is not integrable in $\mathbb{R}^+$.
To deal with this case we introduce a function $g(\cdot; p)$, tuned by a parameter $p > 0$, such that

- $\int_0^\infty g(c; p)dc < \infty$.

This ensures that $f(c;y,w)\cdot g(c;p)$ is integrable for any $p>0$, in fact

$$\begin{aligned}\int_0^\infty f(c;y,w)\cdot g(c;p)dc &= \int_0^\infty \sum_{k=1}^\infty \left(\langle y-w,\theta_k\rangle\right)^2\cdot\exp(-\lambda_k c)\cdot g(c;p)dc\\ &\le \int_0^\infty \sum_{k=1}^\infty \left(\langle y-w,\theta_k\rangle\right)^2\cdot g(c;p)dc\\ &= \left(\int_0^\infty g(c;p)dc\right)\sum_{k=1}^\infty \left(\langle y-w,\theta_k\rangle\right)^2\\ &= \left(\int_0^\infty g(c;p)dc\right) d_e^2(y,w) \;<\; \infty\end{aligned}$$

Without loss of generality we choose g is such that $\int_0^\infty g(c;p)dc = p$. Now, we are able to construct a corresponding distance defined as

$$d_p(y,w) := \sqrt{\int_0^\infty f(c;y,w)\cdot g(c;p)dc} = \sqrt{\int_0^\infty \sum_{k=1}^\infty \left(\langle y-w,\theta_k\rangle\right)^2\cdot\exp(-\lambda_k c)\cdot g(c;p)dc} \tag{2.2}$$

that is finite for any couple of functions y and w. To prove that $d_p(y,w)$ is a distance it's sufficient to note that $d_p(y,w)$ is the $L^2(T)-$distance between the following two elements

$$\widetilde{y} = \sum_{k=0}^\infty \langle y,\theta_k\rangle \left(\int_0^\infty \exp(-\lambda_k c)\cdot g(c;p)dc\right)^{\frac{1}{2}}\cdot\theta_k,$$

$$\widetilde{w} = \sum_{k=0}^\infty \langle w,\theta_k\rangle \left(\int_0^\infty \exp(-\lambda_k c)\cdot g(c;p)dc\right)^{\frac{1}{2}}\cdot\theta_k.$$

where $\widetilde{y}$ and $\widetilde{w}$ are elements of $L^2(T)$ since $\int_0^\infty \exp(-\lambda_k c)\cdot g(c;p)dc \le \int_0^\infty g(c;p)dc < \infty$.

Observe that if the function g is such that for any $p>0$

- $g(c;p)$ is a non increasing and non negative function in c,
- $g(0;p)=1$,

then $f\cdot g$ and f satisfy the same properties, in the sense that $f(0;y,w)\cdot g(0;p) = d^2_{L^2(T)}(y,w)$ and, for any fixed $c\in\mathbb{R}^+$, $\sqrt{f(c;y,w)\cdot g(c;p)}$ represents a distance among y and w dominated by the euclidian distance.

Moreover, for any fixed $c\in(0,\infty)$, we assume that g satisfies the following

- $g(c;p)$ is a non decreasing and non negative function of p.
- $\lim_{p\to\infty} g(c;p) = 1$.

As a consequence, the larger p the greater the distance $d_p(y,w)$. Moreover, even if $d_p(y,w)$ is finite for any couple of functions y and w, when p goes to infinity $d_p(y,w)$ can diverge since it tends to the Mahalanobis distance that can diverge in the infinite dimensional case.

In order to write the distance $d_p(y, w)$ in a more suitable way, let us define

$$h_k(p) \ := \ \int_0^\infty \lambda_k \exp(-\lambda_k c) \cdot g(c; p) dc$$

and so

$$d_p(y, w) = \sqrt{\sum_{k=1}^{\infty} d^2_{M,k}(y, w) \cdot h_k(p)}$$

where $d_{M,k}(y, w) = \sqrt{\left(\langle y - w, \theta_k \rangle\right)^2 / \lambda_k}$ is the term representing the contribution of the Mahalanobis distance along the k^{th} component. We can prove the following results

- $\lim_{k\to\infty} h_k(p) = 0$.
- $\lim_{p\to\infty} h_k(p) = 1$ for any k such that $\lambda_k > 0$.

In fact,

$$\begin{aligned}
\lim_{k\to\infty} h_k(p) &= \lim_{k\to\infty} \int_0^\infty \lambda_k \exp(-\lambda_k c) \cdot g(c; p) dc \\
&= \left(\lim_{k\to\infty} \lambda_k\right) \cdot \left(\int_0^\infty \left(\lim_{k\to\infty} \exp(-\lambda_k c)\right) \cdot g(c; p) dc\right) \\
&= \left(\lim_{k\to\infty} \lambda_k\right) \cdot \left(\int_0^\infty g(c; p) dc\right) \ = \ 0
\end{aligned}$$

and

$$\begin{aligned}
\lim_{p\to\infty} h_k(p) &= \lim_{p\to\infty} \int_0^\infty \lambda_k \exp(-\lambda_k c) \cdot g(c; p) dc \\
&= \int_0^\infty \lambda_k \exp(-\lambda_k c) \cdot \left(\lim_{p\to\infty} g(c; p)\right) dc \\
&= \int_0^\infty \lambda_k \exp(-\lambda_k c) dc \ = \ 1.
\end{aligned}$$

There are many different ways to choose g. Some meaningful examples are:

- $g(c; p) = \mathbf{1}_{\{c \le p\}}$. In this case $h_k(p) = (1 - \exp(-\lambda_k p))$.
- $g(c; p) = \exp(-c/p)$. In this case $h_k(p) = \lambda_k/(\lambda_k + 1/p)$.

3 Inference on the mean of a Gaussian process with known covariance function

The main aim of this paper is to construct testing procedures on the means of Gaussian processes and to discuss their statistical properties. In particular, we propose critical regions based on the generalized distance d_p presented in Section 2. First, the covariance structure is assumed to be known, then in Section 4 we extend these tests to a more general situation where the covariance function is estimated from data.

Let $X_1, .., X_L$ be $L \ge 1$ Gaussian processes with probability laws $P_{X_1}, .., P_{X_L}$, and denote with $m_1, ..., m_L \in L^2(T)$ the corresponding means. Assume that

$P_{X_1}, .., P_{X_L}$ have the same covariance function v, and denote with $(\lambda_k)_k$ the ordered eigenvalues of v and with $(\theta_k)_k$ the associated eigenfunctions. For any $l = 1, .., L$, let $X_{1,l}, ..., X_{n_l,l}$ be n_l i.i.d. realizations of P_{X_l}, and denote with $\bar{X}_l$ the pointwise sample mean: $(X_{1,l} + ... + X_{n_l,l})/n_l$. In this section, we propose critical regions associated to the following inferential problems:

(a) testing the mean of a Gaussian process against an arbitrary function in $L^2(T)$

(b) comparing the means of two Gaussian processes with the same covariance function

Part (a):
Fix $l = 1, .., L$, $m_0 \in L^2(T)$ and consider the following hypothesis test

$$H_0 : \ m_l = m_0 \qquad vs \qquad H_1 : \ m_l \neq m_0. \tag{3.1}$$

To construct a critical region of level α for test (3.1), consider the Karhunen-Loève decomposition of $X_{i,l}$, $i = 1, .., n_l$,

$$X_{i,l}(t) \ = \ m_l(t) \ + \ \sum_{k=1}^{\infty} Z_{ki,l} \sqrt{\lambda_k} \theta_k(t),$$

where $(Z_{ki,l})_{k=1}^{\infty}$ is a sequence of independent standard normal variables, since P_{X_l} is Gaussian. We have that

$$\left(\sqrt{n_l} \cdot \frac{\langle \bar{X}_l - m_l, \theta_k \rangle}{\sqrt{\lambda_k}} \right)_k \ = \ \left(\frac{1}{\sqrt{n_l}} \sum_{i=1}^{n_l} Z_{ki,l} \right)_k \ i.i.d. \ \sim \ \mathcal{N}(0,1) \, .$$

Hence, the sequence $(n_l \cdot d^2_{M,k}(\bar{X}_l, m_l))_k$ is composed by i.i.d. chi-squared random variables with 1 degree of freedom, so that

$$n_l \cdot d^2_p(\bar{X}_l, m_l) \ = \ n_l \cdot \sum_{k=1}^{\infty} d^2_{M,k}(\bar{X}_l, m_l) h_k(p) \ \sim \ \sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$$

where $(\chi^2_{1,k})_{k=1}^{\infty}$ is a sequence of i.i.d. chi-squared variables with 1 d.f. Let us denote as $\xi^2_{\alpha,p}$ the $1-\alpha$ quantile of the distribution of $\sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$. Then, we construct the following critical region of level α

$$R^1_{\alpha} = \left\{ n_l \cdot d^2_p(\bar{X}_l, m_0) > \xi^2_{\alpha,p} \right\} \tag{3.2}$$

for any $p > 0$ and $n_l \in \mathbb{N}$. The quantiles $\xi^2_{\alpha,p}$ are obtained from the distribution of $\sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$ computed in simulation. Note that the probability law of $\sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$ depends on the whole sequence $(\lambda_k)_k$, on the choice of g and on the parameter p.
Let us investigate the power of test (3.1) based on the critical region (3.2). First, note that under the alternative hypothesis

$$d^2_{M,k}(\bar{X}_l, m_0) \ = \ \left(\frac{\langle \bar{X}_l - m_0, \theta_k \rangle}{\sqrt{\lambda_k}} \right)^2 \ = \ \left(\frac{\langle \bar{X}_l - m_l, \theta_k \rangle}{\sqrt{\lambda_k}} + \frac{\langle m_l - m_0, \theta_k \rangle}{\sqrt{\lambda_k}} \right)^2 \ \sim \ \chi^2_{1,k}(\nu_k)$$

where $\chi^2_{1,k}(\nu_k)$ is a non-central chi-squared with 1 d.f. and $\nu_k = d^2_{M,k}(m_l, m_0)$ is the non centrality parameter. Hence, we have that

$$n_l \cdot d^2_p(\bar{X}_l, m_0) \ \sim \ \sum_{k=1}^{\infty} \chi^2_{1,k}(\nu_k) \, h_k(p), \qquad \nu_k = n_l \cdot d^2_{M,k}(m_l, m_0)$$

where $(\chi^2_{1,k}(\nu_k))_{k=1}^{\infty}$ are independent. Then, the power of (3.2) can be obtained as follows

$$\begin{aligned}\beta_1 &= P_{m_l \neq m_0}\left(R^1_\alpha\right) = P_{m_l \neq m_0}\left(n_l \cdot d_p^2(\bar{X}_l, m_0) > \xi^2_{\alpha,p}\right) \\ &= P_{m_l \neq m_0}\left(\sum_{k=1}^{\infty} \chi^2_{1,k}\left(\nu_k\right) h_k(p) > \xi^2_{\alpha,p}\right)\end{aligned}$$

with $\nu_k = n_l \cdot d^2_{M,k}(m_l, m_0)$. Note that the power tends to one as n_l increases.

Remark 3.1 *It is worth to note that this test doesn't have the problem of a low power when m_l and m_0 are arbitrarily distant in $L^2(T)$, which can occur when the inference is computed only with the first components. To see this, consider the quantity $d_p^2(\bar{X}_l, m_0)$ used to compute the power β_1 of the test* (3.2)*. It is possible to show that*

$$\mathbb{E}\left[d_p^2(\bar{X}_l, m_0)\right] = \sum_{k=1}^{\infty}\left[\frac{1}{n_l} + \left(\frac{\langle m_l - m_0, \theta_k\rangle}{\sqrt{\lambda_k}}\right)^2\right] h_k(p) \geq C_1 \cdot d^2_{L^2}(m_l, m_0),$$

$$\mathbb{V}ar\left(d_p^2(\bar{X}_l, m_0)\right) = \sum_{k=1}^{\infty}\left[\frac{2}{n_l^2} + \frac{4}{n_l}\left(\frac{\langle m_l - m_0, \theta_k\rangle}{\sqrt{\lambda_k}}\right)^2\right] h_k(p)^2 \leq C_2 \cdot (1 + p \cdot d^2_{L^2}(m_l, m_0)),$$

where $C_1, C_2 > 0$ are constants independent of m_l and m_0.
From this, we have that $\mathbb{E}\left[d_p^2(\bar{X}_l, m_0)\right] / \sqrt{\mathbb{V}ar\left(d_p^2(\bar{X}_l, m_0)\right)}$ goes to infinity as $d^2_{L^2}(m_l, m_0)$ diverges. Hence, the power β_1 of the test (3.2) *tends to one when $d^2_{L^2}(m_l, m_0)$ increases.*

Part (b):
Fix $l_1, l_2 = 1, .., L$ $(L \geq 2)$, $l_1 \neq l_2$ and consider the following hypothesis test

$$H_0 : \; m_{l_1} = m_{l_2} \qquad vs \qquad H_1 : \; m_{l_1} \neq m_{l_2}. \tag{3.3}$$

The assumption that X_{l_1} and X_{l_2} have the same covariance function can be tested using various inferential procedures presented in literature (see for example Benko *et al.* (2009), Panaretos *et al.* (2010), Fremdt *et al.* (2013), Pigoli *et al.* (2014)). Under the null hypothesis, using the Karunen-Loève decomposition of $\bar{X}_{l_1}$ and $\bar{X}_{l_2}$ we have that

$$\left(\frac{\langle \bar{X}_{l_1} - \bar{X}_{l_2}, \theta_k\rangle}{\sqrt{\lambda_k}}\right)_k \; i.i.d. \; \sim \; \mathcal{N}\left(0, \frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)$$

and then

$$\left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d_p^2(\bar{X}_{l_1}, \bar{X}_{l_2}) \; \sim \; \sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$$

where $(\chi^2_{1,k})_{k=1}^{\infty}$ are all independent. As a consequence, the following critical region

$$R^2_\alpha = \left\{\left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d_p^2(\bar{X}_{l_1}, \bar{X}_{l_2}) > \xi^2_{\alpha,p}\right\} \tag{3.4}$$

is of level α for any $p > 0$ and $n_{l_1}, n_{l_2} \in \mathbb{N}$.
Let us investigate the power of test (3.4). Following similar arguments used in case (a), we obtain that

$$\left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d_p^2(\bar{X}_{l_1}, \bar{X}_{l_2}) \sim \sum_{k=1}^{\infty} \chi_{1,k}^2(\nu_k)\, h_k(p), \quad \nu_k = \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d_{M,k}^2(m_{l_1}, m_{l_2})$$

where $(\chi_{1,k}^2(\nu_k))_{k=1}^{\infty}$ are independent. Then, the power of (3.4) can be obtained as follows

$$\begin{aligned}\beta_2 &= P_{m_{l_1} \neq m_{l_2}}\left(R_\alpha^2\right) = P_{m_{l_1} \neq m_{l_2}}\left(\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \cdot d_p^2(\bar{X}_{l_1}, \bar{X}_{l_2}) > \xi_{\alpha,p}^2\right) \\ &= P_{m_{l_1} \neq m_{l_2}}\left(\sum_{k=1}^{\infty} \chi_{1,k}^2(\nu_k)\, h_k(p) > \xi_{\alpha,p}^2\right)\end{aligned}$$

with $\nu_k = \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d_{M,k}^2(m_{l_1}, m_{l_2})$. Note that the power tends to one when n_{l_1} and n_{l_2} go to infinity.

4 Inference on the mean of a Gaussian process with unknown covariance function

In this section, we extend the inferential procedures presented in Section 3 to the case of unknown covariance structure. We propose tests similar to (3.2) and (3.4), where the covariance operator and related eigenvalues and eigenfunctions are estimated from data. For any $l = 1, .., L$, let us introduce

$$\widehat{v}_{l,n_l} := \frac{1}{n_l - 1} \sum_{i=1}^{n_l} \left(X_{i,l}(s) - \bar{X}_l(s)\right)\left(X_{i,l}(t) - \bar{X}_l(t)\right).$$

the estimator of v computed using the n_l realizations of P_{X_l}: $X_{1,l}, ..., X_{n_l,l}$. Then, call $N = n_1 + ... + n_L$ the total number of realizations from $X_1, ..., X_L$ and define

$$\widehat{v}_N := \frac{1}{N - L} \sum_{l=1}^{L} (n_l - 1)\widehat{v}_{l,n_l} \tag{4.1}$$

the pooled estimator of v computed using all data. Let us denote with $(\widehat{\lambda}_k)_k$ the ordered eigenvalues of $\widehat{v}_N$ and $(\widehat{\theta}_k)_k$ the associated eigenfunctions. Naturally, $\widehat{\lambda}_k = 0\ \forall k \geq N$, so the eigenfunctions $(\widehat{\theta}_N, \widehat{\theta}_{N+1}, ..)$ can be arbitrary chosen such that $(\widehat{\theta}_k)_k$ is an orthonormal basis of $L^2(T)$.

First, it is worth to highlight an important point concerning all the tests presented in this section. The inferential procedures proposed for the case of unknown covariance function are asymptotic, in the sense that the nominal level of the tests is achieved when the size of samples used to estimate v is large. This could be different from the number of data involved in the estimation of the means to be tested. In fact, all the asymptotic results hold for $N \to \infty$, which does not imply the divergence of all the sample sizes $n_1, .., n_L$ (when $L \geq 1$). Naturally, since $N = n_1 + ... + n_L$, at least one among $n_1, .., n_L$ must goes to infinity if $N \to \infty$,

but this is not necessarily the size of the sample drawn from the processes used to estimate the means considered in the hypothesis test.

Here, the critical regions will be constructed with statistics based on estimators of the generalized Mahalanobis distance $\widehat{d}_{p,N}$, since the covariance structure, which is required to compute d_p, is supposed unknown in this section. The estimator of d_p based on the covariance estimator $\widehat{v}_N$ is defined as follows

$$\widehat{d}^2_{p,N}(y,w) \;:=\; \sum_{k=1}^{N-1} \widehat{d}^2_{M,k}(y,w)\cdot \widehat{h}_k(p) \;+\; p\sum_{k=N}^{\infty}\Big(\langle y-w,\widehat{\theta}_k\rangle\Big)^2, \tag{4.2}$$

with $y, w \in L^2(T)$. In (4.2), $\widehat{d}^2_{M,k}(\cdot,\cdot)$ and $\widehat{h}_k(p)$ indicate the quantities $d^2_{M,k}(\cdot,\cdot)$ and $h_k(p)$, with $(\lambda_k)_k$ and $(\theta_k)_k$ replaced by $(\widehat{\lambda}_k)_k$ and $(\widehat{\theta}_k)_k$. Comparing the definition of $\widehat{d}_{p,N}$ in (4.2) and d_p in (2.2), we note how the first $N-1$ components are similar, while the terms $k \geq N$ are different because $\widehat{\lambda}_k = 0\ \forall k \geq N$ and so $\widehat{d}_{M,k}$ would be undefined for $k \geq N$. Then, in (4.2) we have introduced a correction in order to make the estimate $\widehat{d}_{p,N}$ as close as possible to d_p. In particular, since $\lambda_k^{-1}h_k(p) \to p$ as $\lambda_k \to 0$, then in (4.2) we have redefined $\widehat{d}_{M,k}\widehat{h}_k(p) := p\cdot\Big(\langle y-w,\widehat{\theta}_k\rangle\Big)^2$ for any $k \geq N$.

Now we study the asymptotic properties of the covariance estimator in order to construct tests based on the distance $\widehat{d}_{p,N}$ with the same structure of (3.2) and (3.4). First, we need the following auxiliar result:

Theorem 4.1 *For any $N, n_0 \in \mathbb{N}$, let $j_N = (j_{N,1},...,j_{N,n_0})$ be a vector of integers ($j_{N,1},..,j_{N,n_0} \in \mathbb{N}$) and let $(Y_{j_N}, W_{j_N})_N$ be a couple of stochastic processes, independent of the sequence $(\widehat{v}_k)_k$, and such that*

$$\sup_{N\geq 1}\mathbb{E}[\|Y_{j_N} - W_{j_N}\|^2] \;<\; \infty. \tag{4.3}$$

Then, we have that

$$\mathbb{E}\left[|\widehat{d}^2_{p,N}(Y_{j_N},W_{j_N}) - d^2_p(Y_{j_N},W_{j_N})|\right] \;\to_N\; 0. \tag{4.4}$$

The proof of Theorem 4.1 is reported in Appendix A. This result is essential to show through Slutsky's Theorem the statistics based on $\widehat{d}_{p,N}$ and those based on d_p converge to an asymptotic distribution. We highlight this in the following result

Corollary 4.2 *For any $l = 1,..,L$ and for any $m_0 \in L^2(T)$ we have that*

$$n_l\cdot\widehat{d}^2_{p,N}(\bar{X}_l, m_0) \;\xrightarrow{D}_N\; \sum_{k=1}^{\infty}\chi^2_{1,k}(\nu_k)h_k(p) \tag{4.5}$$

where $\nu_k = n_l\cdot d^2_{M,k}(m_l,m_0)$. Moreover, when $L \geq 2$, for any $l_1, l_2 = 1,..,L$, we have that

$$\left(\frac{1}{n_{l_1}}+\frac{1}{n_{l_2}}\right)^{-1}\cdot\widehat{d}^2_{p,N}(\bar{X}_{l_1},\bar{X}_{l_2}) \;\xrightarrow{D}_N\; \sum_{k=1}^{\infty}\chi^2_{1,k}(\nu_k)h_k(p), \tag{4.6}$$

where $\nu_k = \left(\frac{1}{n_{l_1}}+\frac{1}{n_{l_2}}\right)^{-1}\cdot d^2_{M,k}(m_{l_1},m_{l_2})$.

Proof. First, consider equation (4.5). Let us apply Theorem 4.1 with $j_N = n_l$, $Y_{j_N} = \sqrt{n_l} \cdot \bar{X}_l$ and $W_{j_N} = \sqrt{n_l} \cdot m_0$ to obtain

$$n_l \cdot \mathbb{E}\left[|\widehat{d}^2_{p,N}(\bar{X}_l, m_0) - d^2_p(\bar{X}_l, m_0)|\right] \ \to_N \ 0.$$

Then, since in Section 3 we showed that

$$n_l \cdot d^2_p(\bar{X}_l, m_0) \ \xrightarrow{D}_N \ \sum_{k=1}^{\infty} \chi^2_{1,k}(\nu_k) h_k(p)$$

with $\nu_k = n_l \cdot d^2_{M,k}(m_l, m_0)$, the result (4.5) is given by Slutsky's Theorem. To prove equation (4.6) we follow the same arguments used in first part of the proof, with $j_N = (n_{l_1}, n_{l_2})^T$, $Y_{j_N} = \bar{X}_{l_1} \cdot \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1/2}$ and $W_{j_N} = \bar{X}_{l_2} \cdot \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1/2}$. ■

Corollary 4.2 ensures us that, for hypothesis tests (3.1) and (3.3), the critical regions based on covariance estimator have the same structure of those in (3.2) and (3.4) where the covariance function is assumed to be known. However, the asymptotic distribution $\sum_{k=1}^{\infty} \chi^2_{1,k}(\nu_k) h_k(p)$ depends on the eigenvalues of v, which are unknown here. Then, to compute the tests we need this further asymptotic result

Theorem 4.3 *Let $(\chi^2_{1,k})_k$ be a sequence of i.i.d. chi-squared with 1 d.f. independent of $\widehat{v}_N$. Let $\widehat{\xi}^2_{\alpha,p}$ be the $1-\alpha$ quantile of the conditional distribution of $\sum_{k=1}^{\infty} \chi^2_{1,k} \widehat{h}_k(p)$ given $(\widehat{\lambda}_k)_k$. Then, we have that*

$$\widehat{\xi}^2_{\alpha,p} \ \xrightarrow{p}_N \ \xi^2_{\alpha,p} \tag{4.7}$$

The proof of Theorem 4.3 is reported in Appendix B.

Finally, we can use Corollary 4.2 and Theorem 4.3 to construct the critical regions for hypothesis tests (3.1) and (3.3).

Part (a):
Consider the hypothesis tests (3.1). From Slustsky's Theorem, equations (4.5) and (4.7), we have that the following critical region is asymptotically in N of level α:

$$R^3_\alpha \ = \ \left\{ \ n_l \cdot \widehat{d}^2_{p,N}(\bar{X}_l, m_l) > \widehat{\xi}^2_{\alpha,p} \ \right\}. \tag{4.8}$$

Following the same arguments, we can compute the power of test (4.8) as follows

$$\begin{aligned} \beta_3 \qquad &= P_{m_l \neq m_0}\left(R^3_\alpha\right) \ = \ P_{m_l \neq m_0}\left(n_l \cdot d^2_p(\bar{X}_l, m_0) > \widehat{\xi}^2_{\alpha,p}\right) \\ &\to_N \ P_{m_l \neq m_0}\left(\sum_{k=1}^{\infty} \chi^2_{1,k}\left(\nu_k\right) h_k(p) > \xi^2_{\alpha,p}\right) \end{aligned}$$

with $\nu_k = n_l \cdot d^2_{M,k}(m_l, m_0)$.

Part (b):
Assume $L \geq 2$ and consider the hypothesis tests (3.1). From Slustsky's Theorem,

equations (4.6) and (4.7), we have that the following critical region is asymptotically in N of level α:

$$R^4_\alpha \;=\; \left\{ \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot \widehat{d}^2_{p,N}(\bar{X}_{l_1}, \bar{X}_{l_2}) > \widehat{\xi}^2_{\alpha,p} \right\} \tag{4.9}$$

Following the same arguments, we can derive the power of test (4.9) as follows

$$\begin{aligned} \beta_4 \qquad &= P_{m_{l_1} \neq m_{l_2}} \left(R^4_\alpha\right) \;=\; P_{m_{l_1} \neq m_{l_2}} \left(n_l \cdot d^2_p(\bar{X}_l, m_0) > \widehat{\xi}^2_{\alpha,p}\right) \\ &\to_N \; P_{m_{l_1} \neq m_{l_2}} \left(\sum_{k=1}^{\infty} \chi^2_{1,k}\left(\nu_k\right) h_k(p) > \xi^2_{\alpha,p}\right) \end{aligned}$$

with $\nu_k = \left(\frac{1}{n_{l_1}} + \frac{1}{n_{l_2}}\right)^{-1} \cdot d^2_{M,k}(m_{l_1}, m_{l_2})$.

5 Inferential Properties and Simulation Studies

In this section, we highlight some interesting properties concerning the tests presented in Section 3 and 4. In particular, we mainly focus on three aspects:

(1) general properties of the tests power function;

(2) dependence of the test power function on the choice of the parameter p;

(3) asymptotic properties of the tests with unknown covariance function.

The discussion is supported and guided by same simulation studies.

5.1 Simulations on the power function

The first goal is to highlight some general properties of the tests presented in this paper. Without loss of generality, the simulations realized for this task only concern the case $H_0 : m_1 = m_0$ vs $H_1 : m_1 \neq m_0$, with the critical region R^1_α in (3.2). Consider a Gaussian process in $L^2([0,1])$ with probability law P_{X_1}. The mean function $m_1(t)$ is supposed to be unknown, while the covariance functions $v(s,t)$ is defined from the sequences of its eigenvalues $(\lambda_k)_k$, $\lambda_k = (k+1)^{-4}$, and its eigenfunctions: $\theta_1(t) = \mathbf{1}_{\{t\in[0,1]\}}$ and

$$\theta_{2k}(t) = \sqrt{2}\sin(k\cdot 2\pi t)\mathbf{1}_{\{t\in[0,1]\}}, \quad \theta_{2k+1}(t) = \sqrt{2}\cos(k\cdot 2\pi t)\mathbf{1}_{\{t\in[0,1]\}}, \quad k = 1, 2, ...$$

Let us consider the hypothesis test (3.1) with $m_0(t) = t(1-t)\mathbf{1}_{\{t\in[0,1]\}}$ as mean function supposed in H_0 and set $\alpha = 0.05$ as significance level. We construct the critical region R^1_α (3.2) with $n = 10$ i.i.d. Gaussian processes with law P_{X_1}, simulated through the Karunen-Loève expansion, using $K = 100$ components and a grid of 500 equispaced points in $T = [0,1]$. The distance d_p used as statistics in R^1_α (3.2) has been realized with the choice of $h_k(p) = (1 - \exp(-\lambda_k p))$, that comes from $g(c;p) = \mathbf{1}_{\{c\leq p\}}$. In this first part of the section the role of p is not important, so let us fix $p = 10^3$. To better understand the power of R^1_α in (3.2), we compute the power for different mean functions in H_1, defined as follows: $m_{1k}(t) = m_0(t) + 0.03 \cdot \theta_k(t)$, for $k = 1, .., 10$. Note that all the possible mean functions m_{1k} have the same L^2-distance from the tested mean m_0: $d_{L^2}(m_{1k}, m_0) = 0.03$

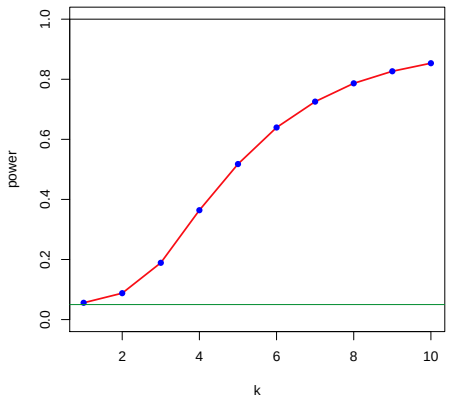


Fig. 1: The empirical power of test (3.2) for $m_{1k}(t) = m_0(t) + 0.03 \cdot \theta_k(t)$, for $k = 1, .., 10$, realized with 10^3 simulations, $n = 10$ and $p = 10^3$.

$\forall k = 1, .., 10$. The empirical power has been computed by realizing 10^3 times the test (3.2), each one using $n = 10$ independent Gaussian processes generated from P_{X_1} (Figure 1). From Figure 1 we note that the power is strictly increasing in k, even if $d_{L^2}(m_{1k}, m_0) = 0.03 \ \forall k = 1, .., 10$. In fact, the test statistics is constructed with the distance d_p, which is able to distinguishes the differences of $m_{1k} - m_0$ along the principal components θ_k and to weight them according to the variability of X along the components λ_k. This is exactly the same idea under the Mahalanobis distance. Naturally, we need to fix a parameter p that rules how much we can distinguish different low variances, since the Mahalanobis distance obtained with $p \to \infty$ is not defined in infinite dimensional space. However, the important thing here is that for any choice of $p > 0$, the test really considers all the infinite components, without any truncation. In the next section will show that the parameter p plays an important statistical role in the inferential properties of the tests and setting p arbitrary large is not in general the right choice for the analysis.

5.2 Simulations on the test power depending on p

The second goal is to discuss how the choice of the parameter p affects the tests presented in the paper. Without loss of generality, we only consider the hypothesis test $H_0 : m_1 = m_0$ vs $H_1 : m_1 \neq m_0$, with the critical region R^1_α expressed in (3.2). The discussion on the role of p for the other critical regions is analogous. In Section 3 we derived the analytic expression of the power of (3.2) as

$$\beta_1 = P\left(\sum_{k=1}^{\infty} \chi^2_{1,k}(\nu_k)\, h_k(p) > \xi^2_{\alpha,p}\right)$$

with $\nu_k = n \cdot d^2_{M,k}(m_1, m_0)$. To better understand how the power function depends on the choice of the parameter p, we compute the power for different values of $p > 0$, when the true mean $m_1(t)$ is one of the following

(a) $m_1(t) = m_0(t)$, the function in the null hypothesis $m_0(t)$ coincides with the true mean $m_1(t) \Rightarrow d_{M,k}(m_1, m_0) = 0 \ \forall k \geq 1$.

(b) $m_1(t) = m_0(t) + \sqrt{\lambda_1} \cdot \theta_1(t)$, $m_0(t)$ and $m_1(t)$ only differs in the first component $\Rightarrow d_{M,1}(m_1, m_0) = 1$, $d_{M,k}(m_1, m_0) = 0 \ \forall k \neq 1$.

(c) $m_1(t) = m_0(t) + \sqrt{\lambda_5} \cdot \theta_5(t)$, $m_0(t)$ and $m_1(t)$ only differs in the 5^{th} component $\Rightarrow d_{M,5}(m_1, m_0) = 1$, $d_{M,k}(m_1, m_0) = 0 \ \forall k \neq 5$.

(d) $m_1(t) = m_0(t) + \sum_{k=10}^{\infty} \sqrt{\lambda_k} \cdot \theta_k(t)$, $m_0(t)$ and $m_1(t)$ differs in all the components but the first 9 $\Rightarrow d_{M,k}(m_1, m_0) = 1 \ \forall k \geq 10$, $d_{M,k}(m_1, m_0) = 0 \ \forall k < 10$.

For each case (a)-(b)-(c)-(d), the power has been computed by replying 10^3 times the test (3.2), each one using $n = 10$ independent Gaussian processes generated from P_{X_1}. In Figure 2 the four cases are separately reported: on the left there is an example of the functional sample $x_1, .., x_{10}$ (yellow lines), the real mean m_1 (red line) and the tested mean m_0 (blue line), while on the right we depict the empirical power of R^1_α (3.2) (blue lines) with $p \in \{10^{-2}, 10^{-1}, ..., 10^8\}$. The dotted orange curves represent the same procedure realized with $n = 20$ and $n = 50$. The aim of these orange lines is to show that the power increases with the sample size n, for any $p > 0$ and for any mean $m_1 \in H_1$ (see cases (b)-(c)-(d)).

In case (a), the null hypothesis is true, then the proportion of rejection of R^1_α (3.2) is equal to the level $\alpha = 0.05$ for any choice of $p > 0$.

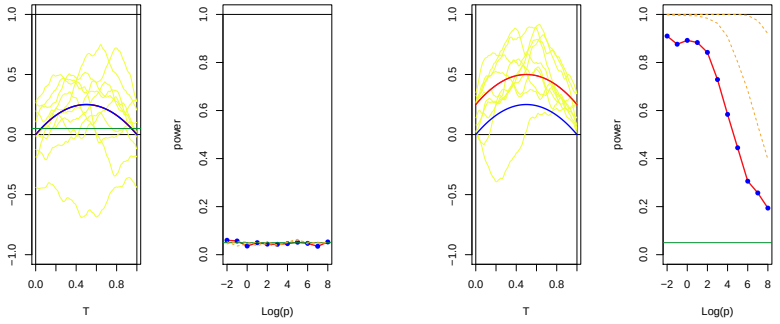
Now, consider case (b), where the first component is the only difference among $m_0(t)$ and $m_1(t)$. When p is small, the term $h_1(p)$ is very relevant with respect to $(h_k(p), k \geq 2)$; then, the non-centrality quantity $d_{M,1}(m_1, m_0) = 1$ multiplied by $h_1(p)$ makes the power higher. However, the more p increases, the more terms $(h_k(p), k = 1, 2, ..)$ become close to one; so, the test statistics $d^2_p(\bar{X}_n, m_0)$ is influenced now by more components and, the first one becomes less relevant; hence, since $m_0(t)$ and $m_1(t)$ only differ in the first component $d_{M,k}(m_1, m_0) = 0 \ \forall k \geq 2$, the power decreases as p increases.

In case (c), the 5^{th} component is the only difference among $m_0(t)$ and $m_1(t)$. Then, for small p we have $h_5(p) \simeq 0$, so the contribute of $d_{M,5}(m_1, m_0) = 1$ is irrelevant and the power is low. As p increases, $h_5(p)$ becomes closer to one and the power grows. However, since $d_{M,k}(m_1, m_0) = 0 \ \forall k > 5$, when p increases a lot, there are too many irrelevant component influencing the test statistics $d^2_p(\bar{X}_n, m_0)$, and the power falls down.

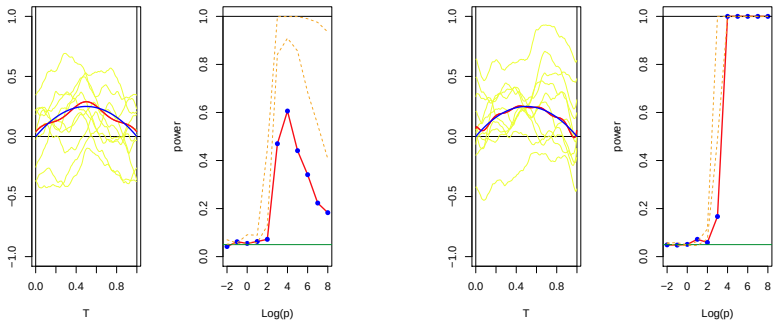
In case (d), $m_0(t)$ and $m_1(t)$ have the same contributions in the first 9 components. When p is small we have that $h_k(p) \simeq 0 \ k \geq 10$, so the test is unable to detect any difference and the power is low. As p increases, some $h_k(p) \ k \geq 10$ become close to one, the statistics $d^2_p(\bar{X}_n, m_0)$ is influenced by $d_{M,k}(m_1, m_0) = 1$ $k \geq 10$ and the power grows. When p goes to infinity, almost all components of the statistics $d^2_p(\bar{X}_n, m_0)$ have a contribution $d_{M,k}(m_1, m_0) = 1$, so that the power tends to one.

5.3 Discussion on the role of p

From Figure 2, we note that the choice of the parameter $p > 0$ somehow determines in which components the test R^1_α (3.2) performs well in detecting possible differences among the supposed mean $m_0(t)$ and true mean $m_1(t)$. In fact, for any given m_0,



(a) $m_1 = m_0$ (b) $m_1 = m_0 + \sqrt{\lambda_1} \cdot \theta_1$



(c) $m_1 = m_0 + \sqrt{\lambda_5} \cdot \theta_5$ (d) $m_1 = m_0 + \sum_{k=10}^{\infty} \sqrt{\lambda_k} \cdot \theta_k$

Fig. 2: For each case $(a) - (b) - (c) - (d)$ we have two figures; on the left: the functional sample $x_1, .., x_{10}$ (yellow lines), the real mean m_1 (red line) and the tested mean m_0 (blue line); on the right: the empirical power of test (3.2) for $p \in \{10^{-2}, 10^{-1}, ..., 10^8\}$ realized with 10^3 simulations, with $n = 10$ (red lines), $n = 20$ and $n = 50$ (orange dotted lines).

the test (3.1) can provide different results according to the value of the parameter p.

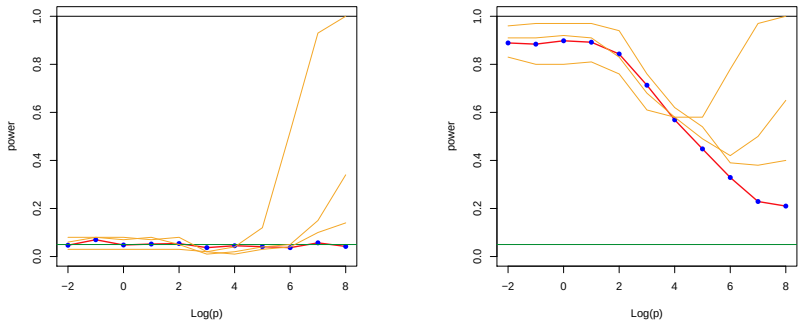
When we set a small p, we obtain a test whose power of detecting differences between m_0 and the true mean m_1 is not indicated to discriminate many different components with their own variability, but is more adapted to look at the shape of the function. In other words, the distance d_p among m_1 and m_0 is closer to the L^2-distance than the Mahalanobis one. This can be a good choice when we are interested in the general closeness among the supposed mean m_0 and the true m_1, regardless the micro-structure's features. For instance, in case (b) m_0 is quite far from m_1 and the test with low p works well. At the contrary, in case (c) and (d) the macro-structure of m_0 and m_1 are not so different and so for low p the power is low. When we set a large p, the test explores much more components taking into account their own variability, since in the distance d_p many weights $h_k(p)/\lambda_k$ are close to $1/\lambda_k$. Hence, the distance d_p among m_1 and m_0 which influences the power is more inspired to the Mahalanobis distance's idea than to the classical L^2-distance. Despite of this good property, there are two aspects that we should be aware of when we set a large p:

- each component is less relevant in the test statistics. For instance, in the cases (b) and (c), $m_0(t)$ and $m_1(t)$ differ for only one component; then, when p is very large that component becomes negligible and the power of the test decreases. The opposite case is (d), in which the power tends to one for large p because the first nine components, which are in favour of H_0, become negligible.
- the test statistics less reflects the macro-structure of the function. For instance, in case (b), the curves $m_0(t)$ and $m_1(t)$ seem very far because the difference among $m_0(t)$ and $m_1(t)$ concerns the first component. However, for large p the test is more interested in looking at the average behaviour of many components and that distance easily visualized in the graphic is forgotten. The opposite case is (d), in which $m_0(t)$ and $m_1(t)$ seem very similar at first sight. However, for large p the test doesn't care about the visualization of the curves and it's more concentrated on the average difference in many components, which leads to reject H_0 with high probability.

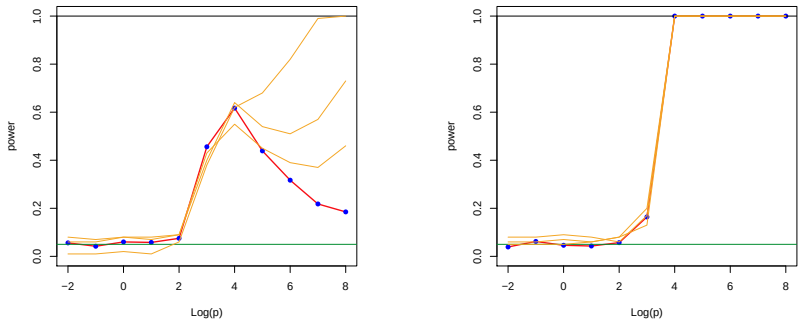
In conclusion, there is not an optimal choice for p: according to the main interest associated to the test, we should opportunely set the parameter p to detect differences related to a specific aim.

5.4 Simulations on tests with unknown covariance function

In this part, we want to to show that the critical regions based on the estimated covariance function present asymptotically the same power of the critical regions with known covariance structure. We remind that the term asymptotic here is referred to the number of data used to estimate the covariance function. We realized simulations for the hypothesis test $H_0 : m_1 = m_0$ vs $H_1 : m_1 \neq m_0$, with the critical region R^3_α expressed in (4.8), since the properties of test aiming at comparing means from two populations are analogous. We adopt the same framework presented in the previous part of this section. In addition to the $n_1 = 10$ processes from P_{X_1} previously defined, let us introduce n_2 i.i.d. Gaussian processes with probability law P_{X_2}, having the same covariance function $v(s,t)$ of P_{X_1}. These $N = n_1 + n_2$



(a) $m_1 = m_0$ (b) $m_1 = m_0 + \sqrt{\lambda_1} \cdot \theta_1$



(c) $m_1 = m_0 + \sqrt{\lambda_5} \cdot \theta_5$ (d) $m_1 = m_0 + \sum_{k=10}^{\infty} \sqrt{\lambda_k} \cdot \theta_k$

Fig. 3: For each case $(a) - (b) - (c) - (d)$ we represent the empirical power of tests (3.2) (red line) and (4.8) (orange lines) for $p \in \{10^{-2}, 10^{-1}, ..., 10^{8}\}$, $n_1 = 10$, $n_2 \in \{100, 500, 1000\}$, realized with 10^3 simulations.

processes provides the covariance estimate $\widehat{v}_N$ as expressed in (4.1). For each $n_2 \in \{100, 500, 1000\}$, we compute the power of R^3_α (4.8) as function of the parameter $p \in \{10^{-2}, 10^{-1}, ..., 10^{8}\}$. This procedure is repeated changing the true mean m_1 of P_{X_1}, as specified in the cases $(a) - (b) - (c) - (d)$ These power functions are represented in Figure 3 with orange dotted lines. The blue lines indicate the power of test R^1_α (3.2) adopted when the covariance v is known. As we can see form Figure 3 the power functions of R^3_α (4.8) tends to the power of R^1_α (3.2) as N goes to infinity. Moreover, this convergence is faster when p is smaller and slower when p is large.

Appendix A

Proof. [Proof of Theorem 4.1] The purpose of the proof is to show that $\forall \epsilon > 0$

$$\lim_{N\to\infty} \mathbb{E}\left[|\widehat{d}^2_{p,N}(Y_{j_N}, W_{j_N}) - d^2_p(Y_{j_N}, W_{j_N})|\right] \;\leq\; \epsilon.$$

First, fix an integer $k_\epsilon \geq 1$ such that

$$2p \sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right] \;<\; \epsilon. \tag{5.1}$$

The existence of k_ϵ is ensured by Condition (4.3). Then, consider the following decomposition

$$\begin{aligned}
&\mathbb{E}\left[|\widehat{d}^2_{p,N}(Y_{j_N}, W_{j_N}) - d^2_p(Y_{j_N}, W_{j_N})|\right] \;\leq\; \sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[d^2_{M,k}(Y_{j_N}, W_{j_N})\right] \cdot h_k(p) \\
&+ \sum_{k=k_\epsilon+1}^{N-1} \mathbb{E}\left[\widehat{d}^2_{M,k}(Y_{j_N}, W_{j_N}) \cdot \widehat{h}_k(p)\right] \;+\; p \sum_{k=N}^{\infty} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k\rangle\right)^2\right] \\
&+ \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[|d^2_{M,k}(Y_{j_N}, W_{j_N}) \cdot h_k(p) - \widehat{d}^2_{M,k}(Y_{j_N}, W_{j_N}) \cdot \widehat{h}_k(p)|\right] \\
&= A_N \;+\; B_N \;+\; C_N.
\end{aligned}$$

First, consider the term A_N. We have that

$$\begin{aligned}
A_N \;&=\; \sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[d^2_{M,k}(Y_{j_N}, W_{j_N})\right] \cdot h_k(p) \;=\; \sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right] \cdot \frac{h_k(p)}{\lambda_k} \\
&\leq p \cdot \left(\sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right]\right) \;<\; \epsilon/2
\end{aligned}$$

where the last passage is due to (5.1).
Second, consider the term B_N. Since $h_x(p) \leq xp \;\forall x > 0$, we have that

$$\begin{aligned}
B_N \;&=\; \sum_{k=k_\epsilon+1}^{N-1} \mathbb{E}\left[\widehat{d}^2_{M,k}(Y_{j_N}, W_{j_N}) \cdot \widehat{h}_k(p)\right] \;+\; p \sum_{k=N}^{\infty} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k\rangle\right)^2\right] \\
&\leq p \cdot \left(\sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k\rangle\right)^2\right]\right) \\
&= p \cdot \left(\mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2\right] - \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k\rangle\right)^2\right]\right) \\
&= p \cdot \left(\sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right]\right) \\
&- p \cdot \left(\sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k\rangle\right)^2 - (\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right]\right)
\end{aligned}$$

and by using the same arguments adopted to treat the term A_N we obtain

$$B_N \quad \leq \epsilon/2 \; + \; p \cdot \left(\sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| \left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k \rangle \right)^2 - (\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \right| \right] \right)$$

Now, consider the second term by defining

$$D_N := \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| \left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k \rangle \right)^2 - (\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \right| \right].$$

and note that

$$\begin{aligned} D_N \quad &= \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| \left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k - \theta_k \rangle + \langle Y_{j_N} - W_{j_N}, \theta_k \rangle \right)^2 - (\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \right| \right] \\ &\leq \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k - \theta_k \rangle \right)^2 + 2|\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k - \theta_k \rangle||\langle Y_{j_N} - W_{j_N}, \theta_k \rangle| \right] \end{aligned}$$

Now, for any $k = 1, ..., k_\epsilon$, by using Cauchy-Shwarz we obtain

$$\leq \mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \|\widehat{\theta}_k - \theta_k\|^2 \right] + 2\sqrt{\mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \|\widehat{\theta}_k - \theta_k\|^2 \right] \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \right]}$$

From $(\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \leq \|Y_{j_N} - W_{j_N}\|^2$ and since $(\widehat{v}_k)_k$ and $(Y_{j_N}, W_{j_N})_N$ are independent we get that

$$\leq \mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \right] \mathbb{E}\left[\|\widehat{\theta}_k - \theta_k\|^2 \right] + 2\mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \right] \sqrt{\mathbb{E}\left[\|\widehat{\theta}_k - \theta_k\|^2 \right]}$$

Then,

$$\begin{aligned} D_N \quad &\leq \mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \right] \sum_{k=1}^{k_\epsilon} \left(\left[\|\widehat{\theta}_k - \theta_k\|^2 \right] + 2\sqrt{\mathbb{E}\left[\|\widehat{\theta}_k - \theta_k\|^2 \right]} \right) \\ &\leq \mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2 \right] k_\epsilon \max_{k=1,..,k_\epsilon} \left\{ \left[\|\widehat{\theta}_k - \theta_k\|^2 \right] + 2\sqrt{\mathbb{E}\left[\|\widehat{\theta}_k - \theta_k\|^2 \right]} \right\} \; \to_N \; 0 \end{aligned}$$

from the consistency of estimated eigenfunctions (see Dauxois *et al.* (1982), Bosq (2000), Horvath and Kokoszka (2012)). This implies that $B_n < \epsilon/2 + o(1)$.
Finally, consider the term C_N. We have that

$$\begin{aligned} C_N \quad &\leq \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| \left(\frac{\langle Y_{j_N} - W_{j_N}, \theta_k \rangle}{\sqrt{\lambda_k}} \right)^2 \cdot h_k(p) - \left(\frac{\langle Y_{j_N} - W_{j_N}, \theta_k \rangle}{\sqrt{\widehat{\lambda}_k}} \right)^2 \cdot \widehat{h}_k(p) \right| \right] \\ &+ \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| \left(\frac{\langle Y_{j_N} - W_{j_N}, \theta_k \rangle}{\sqrt{\widehat{\lambda}_k}} \right)^2 \cdot \widehat{h}_k(p) - \left(\frac{\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k \rangle}{\sqrt{\widehat{\lambda}_k}} \right)^2 \cdot \widehat{h}_k(p) \right| \right] \\ &\leq \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 \left| \frac{h_k(p)}{\lambda_k} - \frac{\widehat{h}_k(p)}{\widehat{\lambda}_k} \right| \right] \\ &+ p \cdot \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[\left| (\langle Y_{j_N} - W_{j_N}, \theta_k \rangle)^2 - \left(\langle Y_{j_N} - W_{j_N}, \widehat{\theta}_k \rangle \right)^2 \right| \right] \; = \; E_N \; + \; D_N \end{aligned}$$

We showed above that $D_N \to_N 0$, then let us consider the term E_n. Since $(Y_{j_N}, W_{j_N})_N$ is independent of $(\widehat{v}_k)_k$, and since the function $x \mapsto h_x(p)/x$ is continuous, we have that

$$\begin{aligned} E_n &\leq \left(\sum_{k=1}^{k_\epsilon} \mathbb{E}\left[(\langle Y_{j_N} - W_{j_N}, \theta_k\rangle)^2\right]\right) \max_{k=1,..,k_\epsilon} \left\{\mathbb{E}\left[\left|\frac{h_k(p)}{\lambda_k} - \frac{\widehat{h}_k(p)}{\widehat{\lambda}_k}\right|\right]\right\} \\ &\leq \sup_{n\geq 1}\left\{\mathbb{E}\left[\|Y_{j_N} - W_{j_N}\|^2\right]\right\} \cdot M_p \max_{k=1,..,k_\epsilon} \left\{\mathbb{E}\left[\left|\lambda_k - \widehat{\lambda}_k\right|\right]\right\} \to_n 0, \end{aligned}$$

from the consistency of estimated eigenvalues (see Dauxois *et al.* (1982), Bosq (2000), Horvath and Kokoszka (2012)), that implies that $C_n = o(1)$.
Putting all together we have shown that

$$\lim_{N\to\infty} \mathbb{E}\left[|\widehat{d}^2_{p,N}(Y_{j_N}, W_{j_N}) - d^2_p(Y_{j_N}, W_{j_N})|\right] \leq \epsilon.$$

for any $\epsilon > 0$, which concludes the proof. ■

Appendix B

Proof. [Proof of Theorem 4.3] To obtain the thesis we should prove that

$$\sum_{k=1}^{\infty} \chi^2_{1,k} \widehat{h}_k(p) \xrightarrow{D} \sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)$$

where $(\chi^2_{1,k})_k$ is a sequence of i.i.d. chi-squared with 1 d.f. independent of $\widehat{v}_N$. In particular, we will show that $\forall \epsilon > 0$

$$\lim_{n\to\infty} \mathbb{E}\left[\left|\left(\sum_{k=1}^{\infty} \chi^2_{1,k} \widehat{h}_k(p)\right) - \left(\sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)\right)\right|\right] \leq \epsilon,$$

so that the convergence holds in L^1. First, fix an integer $k_\epsilon \geq 1$ such that

$$2p \sum_{k=k_\epsilon+1}^{\infty} \lambda_k < \epsilon. \tag{5.2}$$

In fact, note that

$$\begin{aligned} \mathbb{E}\left[\left|\left(\sum_{k=1}^{\infty} \chi^2_{1,k} \widehat{h}_k(p)\right) - \left(\sum_{k=1}^{\infty} \chi^2_{1,k} h_k(p)\right)\right|\right] &\leq \mathbb{E}\left[\sum_{k=1}^{\infty} \chi^2_{1,k} |\widehat{h}_k(p) - h_k(p)|\right] \\ &= \sum_{k=1}^{\infty} \mathbb{E}\left[|\widehat{h}_k(p) - h_k(p)|\right] \end{aligned}$$

since $(\chi^2_{1,k})_k$ is independent of $\widehat{v}_N$. Then, we decompose the series in the following three terms

$$\begin{aligned} &= \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[|\widehat{h}_k(p) - h_k(p)|\right] + \sum_{k=k_\epsilon+1}^{\infty} \mathbb{E}\left[\widehat{h}_k(p)\right] + \sum_{k=k_\epsilon+1}^{\infty} h_k(p) \\ &= A_N + B_N + C_N. \end{aligned}$$

Since the function $x \mapsto h_x(p)$ is continuous, for the first term we have

$$A_N \;=\; \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[|\widehat{h}_k(p) - h_k(p)|\right] \;\leq\; M_p \cdot \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[|\widehat{\lambda}_k - \lambda_k|\right] \;=\; M_p k_\epsilon \cdot \left(\max_{k=1,..,k_\epsilon} \left\{ \mathbb{E}\left[|\widehat{\lambda}_k - \lambda_k|\right] \right\} \right)$$

that tends to zero from the consistency of estimated eigenvalues (see Dauxois *et al.* (1982), Bosq (2000), Horvath and Kokoszka (2012)).
Concerning the term C_N we have

$$C_N \;=\; \sum_{k=k_\epsilon+1}^{\infty} h_k(p) \;=\; p \sum_{k=k_\epsilon+1}^{\infty} \lambda_k \;<\; \frac{\epsilon}{2}$$

from (5.2).
Then, consider the term B_N. We have

$$B_N \;\leq\; p \cdot \mathbb{E}\left[\sum_{k=k_\epsilon+1}^{\infty} \widehat{\lambda}_k \right] \;=\; p \cdot \left(\mathbb{E}\left[\sum_{k=1}^{\infty} \widehat{\lambda}_k \right] \;-\; \mathbb{E}\left[\sum_{k=1}^{k_\epsilon} \widehat{\lambda}_k \right] \right)$$

where the first term can be written as

$$\mathbb{E}\left[\sum_{k=1}^{\infty} \widehat{\lambda}_k \right] \;=\; \mathbb{E}\left[\int \widehat{v}_N(t,s) dt ds \right] \;=\; \int \mathbb{E}\left[\widehat{v}_N(t,s) \right] dt ds \;=\; \sum_{k=1}^{\infty} \lambda_k .$$

Then, putting all together we have

$$\begin{aligned} B_N \quad &\leq p \cdot \left(\sum_{k=1}^{\infty} \lambda_k \;-\; \mathbb{E}\left[\sum_{k=1}^{k_\epsilon} \widehat{\lambda}_k \right] \right) \;=\; p \cdot \left(\sum_{k=k_\epsilon+1}^{\infty} \lambda_k \;-\; \mathbb{E}\left[\sum_{k=1}^{k_\epsilon} (\widehat{\lambda}_k - \lambda_k) \right] \right) \\ &\leq p \cdot \left(\sum_{k=k_\epsilon+1}^{\infty} \lambda_k \;+\; \sum_{k=1}^{k_\epsilon} \mathbb{E}\left[|\widehat{\lambda}_k - \lambda_k| \right] \right) \;<\; \frac{\epsilon}{2} \;+\; p \cdot k_\epsilon \cdot \left(\max_{k=1,..,k_\epsilon} \left\{ \mathbb{E}\left[|\widehat{\lambda}_k - \lambda_k| \right] \right\} \right) \end{aligned}$$

and the last quantity tends to zero from the consistency of estimated eigenvalues (see Dauxois *et al.* (1982), Bosq (2000), Horvath and Kokoszka (2012)). This concludes the proof. ■

References

[1] Benko, M., Hardle, W. and Kneip, A. (2009). Common functional principal components. *Annals of Statistics*, vol. 37, p. 1–34.

[2] Besse, P. and Ramsay, J. (1986). Principal components of sampled functions. *Psychometrika*, vol. 51, p. 285–311.

[3] Bosq, D. (2000). Linear Processes in Function Spaces - Theory and Applications. Springer Series in Statistics, Berlin.

[4] Bunea, F., Ivanescu, A.E. and Wegkamp, M.H. (2011). Adaptive inference for the mean of a Gaussian process in functional data. *Journal of Royal Statistical Society B*, vol. 73, Part. 4, p. 531–558.

[5] Cao, G., Yang, L. and Todem, D. (2012). Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, vol. 24, No. 2, p. 359–377.

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

[7] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, vol. 12, p. 136–154.

[8] Degras, D.A. (2011). Simultaneous Confidence Bands for Nonparametric Regression with Functional Data. *Journal of Nonparametric Statistics*, vol. 24, No. 2, p. 359–377. *Statistica Sinica*, vol. 21, p. 1735–1765.

[9] Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Berlin.

[10] Fremdt, S., Steinebach, J.G., Horvth, L. and Kokoszka, P (2013). Testing the Equality of Covariance Operators in Functional Samples. *Scandinavian Journal of Statistics*, vol. 40, p. 138–152.

[11] Gervini, D. (2006). Free-knot spline smoothing for functional data. *Journal of Royal Statistical Society B*, vol. 68, p. 671–687.

[12] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of Royal Statistical Society B*, vol. 68, p. 109–126.

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

[14] Lacantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T. and Cohen, K.L. (1999). Robust principal component analysis for functional data. *Test*, vol. 8, p. 1–73.

[15] Li, Y. and Hsing, T. (2010). Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data. *Annals of Statistics*, vol. 38, p. 3321–3351.

[16] Luschgy, H. (1991). Testing One-Sided Hypothesis for the Mean of a Gaussian Process. *Metrika*, vol. 38, p. 179–194.

[17] Ma, S., Yang, L. and Carroll, L.J. (2012). A Simultaneous Confidence Band for Sparse Longitudinal Data. *Statistica Sinica*, vol. 22, p. 95–122.

[18] Panaretos, V. M., Kraus, D. and Maddocks, J.H. (2010). Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles. *Journal of the American Statistical Association (Theory and Methods)*, vol. 105, p. 670–682.

[19] Pezzulli, S.D. and Silverman, B. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, vol. 8, p. 1–16.

[20] Pigoli, D., Aston, J.A.D., Dryden, I.L. and Secchi, P. (2014). Distances and inference for covariance operators. *Biometrika*, vol. 101, p. 409–422.

[21] Ramsey, J. and Dalzell, C.J. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society B*, vol. 53, p. 539–572.

[22] Ramsey, J. and Silverman, B. (2002). Applied Functional Data Analysis. New York, Springer.

[23] Ramsey, J. and Silverman, B. (2005). Functional Data Analysis. New York, Springer.

[24] Rice, J.A. and Silverman, B. (1991). Estimating the mean ad covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society B*, vol. 53, p. 233–243.

[25] Ruppert, D., Wand., M. and Carroll, R. (2003). Semiparametric Regression. Cambridge University Press.

[26] Secchi, P., Stamm, A. and Vantini, S. (2013). Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of Hotellings theorem. *Elenctronic Journal of Statistics*, vol. 7, p. 2005–2031.

[27] Silverman, B. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, vol. 24, p. 1–24.

[28] Yao, F., Muller, H.G. and Wang, J.L (2005). Functional data analysis for sparse longitudinal data. *Journal of American Statistical Association*, vol. 100, p. 577–590.

[29] Yao, F. (2007). Asymptotic distributions of nonparametric regression estimators for longitudinal and functional data. *Journal of Multivariate Analysis*, vol. 98, p. 40–56.

[30] Zhang, J.T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, vol. 35, p. 1052–1079.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

39/2014 Ghiglietti, A.; Paganoni, A.M.
Statistical inference for functional data based on a generalization of Mahalanobis distance

38/2014 Shen, H.; Truong, Y.; Zanini, P.
Independent Component Analysis for Spatial Stochastic Processes on a Lattice

37/2014 Giuliani, N.; Mola, A.; Heltai, L.; Formaggia,L.
FEM SUPG stabilisation of mixed isoparametric BEMs: application to linearised free surface flows

36/2014 Abb,A.;Bonaventura,L.; Nini, M.;Restelli,M.;
Anisotropic dynamic models for Large Eddy Simulation of compressible flows with a high order DG method

35/2014 Tricerri, P.; Ded, L.; Deparis, S.; Quarteroni, A.; Robertson, A.M.; Sequeira, A.
Fluid-structure interaction simulations of cerebral arteries modeled by isotropic and anisotropic constitutive laws

34/2014 Antonietti, P.F.; Pacciarini, P.; Quarteroni, A.
A discontinuous Galerkin Reduced Basis Element method for elliptic problems

33/2014 Canuto, C.; Simoncini, V.; Verani, M.
Contraction and optimality properties of an adaptive Legendre-Galerkin method: the multi-dimensional case

32/2014 Agosti, A.; Formaggia, L.;Scotti, A.
Analysis of a model for precipitation and dissolution coupled with a Darcy flux

31/2014 Corno, J.; de Falco, C.; De Gersem, H.; Schps, S.
Isogeometric Simulation of Lorentz Detuning in Superconducting Accelerator Cavities

30/2014 Ferroni, A.; Formaggia, L.; Fumagalli, A.;
Numerical analysis of Darcy problem on surfaces